A COMPREHENSIVE MODERN TREATMENT
OF ANALYTIC FUNCTION THEORY
AND ITS RAMIFICATIONS

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE

VOLUME III

A. I. MARKUSHEVICH

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE

VOLUME III

A. I. MARKUSHEVICH

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE

Volume III

by A. I. MARKUSHEVICH

translated from the Russian by

RICHARD A. SILVERMAN

The present volume in the series SELECTED RUSSIAN PUBLICATIONS IN THE MATHEMATICAL SCIENCES is the third of a three-volume set devoted to analytic function theory and its ramifications. Taken as a whole, the course is unusually detailed, and in addition to covering all the material conventionally found in first courses on the subject, explores a variety of topics rarely encountered at this level. The treatment is selfcontained and rigorous, but at the same time highly motivated. Thus clarity of presentation is never sacrificed to abstractness for its own sake. An abundance of worked-out examples, hundred of relevant problems (many with hints and answers), numerous illustrations and ample bibliographic material combine to make this comprehensive course eminently suitable both for classroom use and self-study. The result is a unique contribution to the serious textbook literature on functions of a complex variable and their applications.

The Author

A. I. MARKUSHEVICH is Professor of Mathematics at Moscow State University and Vice-President, Academy of Pedagogical Sciences, R.S.F.S.R.

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE

Volume III

SELECTED RUSSIAN PUBLICATIONS IN THE MATHEMATICAL SCIENCES

Richard A. Silverman, *Editor*

ABRIKOSOV ET AL. *Methods of Quantum Field Theory in Statistical Physics*
GELFAND & FOMIN *Calculus of Variations*
LEBEDEV *Special Functions and Their Applications*
LEBEDEV, SKALSKAYA & UFLYAND *Problems of Mathematical Physics*
MARKUSHEVICH *Theory of Functions of a Complex Variable, Volume I*
MARKUSHEVICH *Theory of Functions of a Complex Variable, Volume II*
MARKUSHEVICH *Theory of Functions of a Complex Variable, Volume III*
PETROVSKI *Ordinary Differential Equations*
SHILOV *An Introduction to the Theory of Linear Spaces*
SHILOV & GUREVICH *Integral, Measure and Derivative: A Unified Approach*
TOLSTOV *Fourier Series*
WAINSTEIN & ZUBAKOV *Extraction of Signals from Noise*
YAGLOM *An Introduction to the Theory of Stationary Random Functions*

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY. LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA PRIVATE LTD., *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE

Volume III

A. I. MARKUSHEVICH

Professor of Mathematics
Moscow State University

Revised English Edition
Translated and Edited by

Richard A. Silverman

PRENTICE-HALL, INC.

Englewood Cliffs, N.J.

© 1967 by
PRENTICE-HALL, INC.
Englewood Cliffs, N.J.

All rights reserved. No part of this book may be reproduced in any form or by any means without permission in writing from the publisher.

Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

Library of Congress Catalog Card Number 65–12876

Printed in the United States of America

AUTHOR'S PREFACE

The theory of analytic functions of a complex variable is one of the outstanding accomplishments of classical mathematics. The rudiments of the theory had already begun to emerge in the eighteenth century, in the investigations of Euler and d'Alembert, but the full flowering of the subject was only possible in the nineteenth century, as a result of the fundamental researches of Cauchy, Riemann and Weierstrass. Its methods are widely used in many branches of mathematics, ranging from number theory to fluid dynamics, elasticity theory and electrodynamics. Even in our own day, when young scientists are so often preoccupied with more abstract branches of mathematics, the theory of analytic functions remains as significant as ever, both as a treasury of flexible and powerful analytic tools, and as a prototype or model of other theories dealing with generalizations of analytic functions in various directions.

It is the author's pleasant duty to cordially thank Dr. Richard A. Silverman, who translated the book into English, introduced many appropriate changes making the book easier to read, and enhanced the English-language edition with an abundance of problems and exercises.

This book stems from lectures given over the course of many years at Moscow University. In commending it to the attention of its new readership, the author would consider his purpose fulfilled if the present edition served to stimulate the interest of young mathematicians in one of the most remarkable branches of exact science, created by two centuries of joint effort on the part of

outstanding scholars of many nations, and if, at the same time, it represented a contribution by him and the translator towards the strengthening of cultural and scientific cooperation between their two countries.

A. I. M.

TRANSLATOR'S PREFACE

It would be difficult to recall in detail the ways in which the present version departs from the Russian original. Suffice it to say that in keeping with my philosophy of "enlightened translation," I thought it entirely appropriate to interact strongly with the initial text, especially since it lies within the area of my mathematical competence. This has led to a redistribution of the subject matter over three English volumes (the Russian original contains some 700 oversized pages, many in very small print), an overall restyling both of the typography and the art work (in particular, more theorems, examples, remarks, etc. have been explicitly identified as such), an adaptation of the bibliographic materials to suit the needs of those who do not know Russian (a class which presumably includes almost all the readers of this series), and so on. Certainly the most significant change is the addition of over 700 relevant problems, closely paralleling the text and stemming from a variety of sources, the most important of which is the excellent collection by Volkovyski, Lunts and Aramanovich (Moscow, 1961). The entire manuscript has also been read by Dr. T. J. Rivlin, who made many helpful suggestions.

Sections marked with asterisks can be omitted on first reading. However, it is only fair to say that they have been so marked merely because of their relative difficulty, and actually contain some of the most interesting material in the book.

R. A. S.

CONTENTS

PART 1 CONFORMAL MAPPING. APPROXIMATION THEORY, Page 1.

1 CONFORMAL MAPPING: RUDIMENTS, Page 3.

1. Conformal Mapping of Annular Domains, 3.
2. Conformal Mapping of Simply Connected Domains, 8.
3. Basic Properties of Univalent Functions, 14.
Problems, 27.

2 CONFORMAL MAPPING: RAMIFICATIONS, Page 32.

4. Conformal Mapping of Sequences of Domains, 32.
5. Curvilinear Half-Intervals, 42.
6. Accessible Boundary Points, 52.
7. Prime Ends, 57.
8. Boundary Behavior of Conformal Mappings, 66.
Problems, 77.

3 APPROXIMATION BY RATIONAL FUNCTIONS AND POLYNOMIALS, Page 80.

9. Locally Analytic Functions, 80.
10. Functions Meromorphic on a Domain, 84.
11. Runge's Theorem and Related Results, 88.
12. Approximation on Closed Domains, 97.
13. Approximation on Continua, 100.
14. Faber Polynomials, 104.
15. Bernstein's Theorem, 112.

3 APPROXIMATION BY RATIONAL FUNCTIONS AND POLYNOMIALS—Continued

16. Approximation in the Mean, 116.
17. Polynomials Orthogonal on a Domain, 121.
Problems, 129.

PART 2 PERIODIC AND ELLIPTIC FUNCTIONS, Page 133.

4 PERIODIC MEROMORPHIC FUNCTIONS, Page 135.

18. Preliminaries, 135.
19. Periodic Entire Functions. Trigonometric Polynomials, 141.
20. Elliptic Functions, 146.
Problems, 152.

5 ELLIPTIC FUNCTIONS: WEIERSTRASS' THEORY, Page 155.

21. Weierstrass' Elliptic Functions, 155.
22. The Functions $\wp(z \mid \alpha, i\beta)$ and $\wp(z \mid \alpha - i\beta, \alpha + i\beta)$, 162.
23. The Differential Equation for $\wp(z)$, 168.
24. Inversion of Elliptic Integrals, 172.
25. The Functions $\zeta(z)$ and $\sigma(z)$, 178.
26. The Addition Theorem for $\wp(z)$, 183.
27. The Spherical Pendulum, 186.
Problems, 191.

6 ELLIPTIC FUNCTIONS: JACOBI'S THEORY, Page 194.

28. Jacobi's Elliptic Functions, 194.
29. Theta Functions and Their Relation to Elliptic Functions, 198.
30. Infinite Product Expansions of Theta Functions, 208.
Problems, 212.

PART 3 RIEMANN SURFACES. ANALYTIC CONTINUATION, Page 215.

7 RIEMANN SURFACES, Page 217.

31. Topological Preliminaries, 217.
32. Abstract Riemann Surfaces, 221.
33. Triangulations, 227.
34. Interior Mappings, 232.
35. Riemann Covering Surfaces, 237.
36. Regular Analytic Curves, 241.
37. The Riemann Surface of a Meromorphic Function, 245.
38. Examples, 250.
Problems, 254.

8 ANALYTIC CONTINUATION, Page 257.

39. Elements. The Complete Analytic Function, 257.
40. Circular Elements. The Monodromy Theorem, 263.
41. Analytic Continuation in a Star, 272.
42. Singular Points. Generalized Elements and the Analytic Configuration, 276.
43. The Analytic Configuration as a Topological Surface, 287.
44. The Analytic Configuration as a Riemann Surface, 294.
45. Algebraic Functions, 297.
Problems, 308.

9 THE SYMMETRY PRINCIPLE AND ITS APPLICATIONS, Page 315.

46. The Symmetry Principle, 315.
47. More on the Schwarz-Christoffel Transformation, 320.
48. Examples, 332.
49. The Modular Function. Picard's First Theorem, 336.
50. Normal Families of Analytic Functions, 340.
51. Picard's Second Theorem. Julia Directions, 343.
Problems, 347.

BIBLIOGRAPHY, Page 351.

INDEX, Page 355.

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE

Volume III

Part 1

CONFORMAL MAPPING APPROXIMATION THEORY

1

CONFORMAL MAPPING: RUDIMENTS

1. Conformal Mapping of Annular Domains

The reader has already encountered the elementary theory of conformal mapping in the earlier volumes of this course (see e.g., Chap. I.8).[1] We now make a deeper study of the subject, incidentally proving some results cited earlier without proof (see Sec. II.27).

Let $w = f(z)$ be a one-to-one conformal mapping of the first kind (briefly, a *conformal mapping*) carrying a domain G in the z-plane into a domain D in the w-plane. Then a theorem due to Menshov asserts that $f(z)$ is univalent.[2] Moreover, according to Theorem I.6.1, $f(z)$ is a homeomorphism, and hence a necessary condition for the existence of a conformal mapping of G onto D is that G be homeomorphic to D. Thus, in particular, a multiply connected domain cannot be mapped conformally onto a simply connected domain (see Theorem I.6.4, Corollary).

However, the fact that G is homeomorphic to D is not a sufficient condition for the existence of a conformal mapping of G onto D. For example, let

[1] We refer to A. I. Markushevich, *Theory of Functions of a Complex Variable, Volumes I, II* (translated by R. A. Silverman), Prentice-Hall, Inc., Englewood Cliffs, N. J. (1965). These two books will henceforth be called simply Volume I and Volume II. Theorems, chapters, sections, formulas, problems, figures and page numbers cited from Volumes I and II will be indicated by the prefixes I and II, respectively. Thus Chap. I.8 means Chapter 8 of Volume I, Sec. II.27 means Section 27 of Volume II, and so on.

[2] See D. E. Menshov, *Les Conditions de Monogénéité*, Actualités scientifiques et industrielles, No. 329, Hermann et Cie., Paris (1936), p. 39 ff. Concerning the terminology, cf. footnote 17, p. II.164.

G be the finite z-plane, and let D be the disk $|w| < R < \infty$. If there were a function $f(z)$ mapping G conformally onto D, then $f(z)$ would be single-valued, analytic and bounded in the whole plane. But then $f(z) \equiv \text{const}$, according to Liouville's theorem (Theorem I.16.8), contrary to the assumption that $f(z)$ is one-to-one. Thus there does not exist a conformal mapping of the finite plane G onto the disk D (of finite radius). On the other hand, G and D are homeomorphic. In fact, the function

$$w = f(z) = \begin{cases} \dfrac{2Rz}{\pi|z|} \arctan |z| & \text{for } z \neq 0, \\ 0 & \text{for } z = 0 \end{cases}$$

is one of an infinite family of homeomorphisms of G onto D.

Remark. Actually, the finite plane can be mapped conformally only onto itself or onto the extended plane minus a single finite point. This follows from the fact that, according to Theorem II.4.3, every function univalent in the whole finite plane must be a Möbius transformation.[3]

Our first theorem shows that in general, one domain cannot be mapped conformally onto another unless certain conditions are satisfied:

THEOREM 1.1. *The annulus* $G: 0 < r_1 < |z| < r_2 < \infty$ *can be mapped conformally onto the annulus* $D: 0 < R_1 < |w| < R_2 < \infty$ *if and only if the two annuli are similar, i.e.,*

$$\frac{R_2}{R_1} = \frac{r_2}{r_1}. \tag{1.1}$$

If this condition is satisfied, the function $w = f(z)$ *mapping* G *onto* D *is a Möbius transformation.*

Proof. The sufficiency is obvious, since if (1.1) holds, the entire linear transformation

$$w = f(z) = \frac{R_1}{r_1} z = \frac{R_2}{r_2} z$$

maps G onto D. To prove the necessity, suppose the function $w = f(z)$, with inverse $z = \varphi(w)$, maps G onto D, and let Γ_ρ be the image of the circle $\gamma_\rho: z = \rho e^{i\theta}$ $(0 \leqslant \theta \leqslant 2\pi,\ r_1 < \rho < r_2)$ under the mapping $w = f(z)$. Then Γ_ρ is a closed Jordan curve, with equation

$$w = f(\rho e^{i\theta}) \qquad (0 \leqslant \theta \leqslant 2\pi).$$

Clearly

$$1 = \frac{1}{2\pi i} \int_{\gamma_\rho} \frac{dz}{z} = \pm \frac{1}{2\pi i} \int_{\Gamma_\rho} \frac{\varphi'(w)}{\varphi(w)}\, dw = \pm \frac{1}{2\pi} \operatorname*{Var}_{w \in \Gamma_\rho} \operatorname{Arg} \varphi(w), \tag{1.2}$$

[3] As usual, entire linear transformations are regarded as Möbius transformations of a particularly simple kind (with ∞ as a fixed point).

where in evaluating the integrals, both curves γ_ρ and Γ_ρ are traversed in the positive direction, and the plus or minus sign is chosen depending on whether the point $w = f(z)$ traverses the curve Γ_ρ in the same direction as the point z traverses the curve γ_ρ, or in the opposite direction. It follows that the origin $w = 0$ lies inside Γ_ρ, and moreover that the integral

$$\frac{1}{2\pi i}\int_{\Gamma_\rho}\frac{\varphi'(w)}{\varphi(w)}\,dw, \tag{1.3}$$

which is a continuous function of ρ, takes only integral values (in fact, ± 1). Therefore the integral (1.3) has a constant value, i.e., the sign chosen in (1.2) does not depend on ρ. In other words, the point $w = f(z)$ either traverses Γ_ρ in the same direction as the point z traverses γ_ρ for all ρ in the interval (r_1, r_2), or else w traverses Γ_ρ in the opposite direction for all such ρ.

Assuming temporarily that γ_ρ and Γ_ρ are traversed in the same direction, we now calculate the area S_ρ bounded by Γ_ρ. According to a familiar formula from calculus,[4] we have

$$S_\rho = \frac{1}{2}\int_{\Gamma_\rho}(u\,dv - v\,du) > 0, \tag{1.4}$$

where $w = u + iv$ and Γ_ρ is traversed in the positive direction. Formula (1.4) can be written as

$$S_\rho = \frac{1}{2}\int_{\Gamma_\rho}\operatorname{Im}\{(u - iv)(du + i\,dv)\} = \frac{1}{2}\int_0^{2\pi}\operatorname{Im}\left\{\overline{f(\rho e^{i\theta})}\,\frac{\partial f(\rho e^{i\theta})}{\partial\theta}\right\}d\theta,$$

where the sign of the second integral is the same as that of the first, since increasing θ from 0 to 2π corresponds to traversing Γ_ρ in the positive direction. Expanding $f(\rho e^{i\theta})$ in Laurent series, we obtain

$$\begin{aligned} f(\rho e^{i\theta}) &= \sum_{n=-\infty}^{\infty} a_n \rho^n e^{in\theta}, \\ \frac{\partial f(\rho e^{i\theta})}{\partial\theta} &= \sum_{n=-\infty}^{\infty} ina_n \rho^n e^{in\theta}. \end{aligned} \tag{1.5}$$

Since these expansions are uniformly convergent in θ (for fixed ρ), we have

$$\begin{aligned} S_\rho &= \frac{1}{2}\int_0^{2\pi}\operatorname{Im}\left\{\sum_{m=-\infty}^{\infty}\bar{a}_m\rho^m e^{-im\theta}\sum_{n=-\infty}^{\infty}ina_n\rho^n e^{in\theta}\right\}d\theta \\ &= \operatorname{Im}\left\{\frac{1}{2}\sum_{n=-\infty}^{\infty}in|a_n|^2\rho^{2n}2\pi\right\} = \pi\sum_{n=-\infty}^{\infty}n|a_n|^2\rho^{2n} \\ &= \pi\sum_{n=1}^{\infty}n|a_n|^2\rho^{2n} - \pi\sum_{n=1}^{\infty}n|a_{-n}|^2\rho^{-2n} > 0. \end{aligned}$$

[4] See e.g., D. V. Widder, *Advanced Calculus*, second edition, Prentice-Hall, Inc., Englewood Cliffs, N. J. (1961), p. 225.

Now, since Γ_ρ lies inside the circle $|w| = R_2$, while the circle $|w| = R_1$ lies inside Γ_ρ, for all ρ in (r_1, r_2), we have the inequality

$$\pi R_1^2 < \pi \sum_{n=-\infty}^{\infty} n|a_n|^2\rho^{2n} < \pi R_2^2 \qquad (r_1 < \rho < r_2). \tag{1.6}$$

It follows that the series

$$\sum_{n=-\infty}^{\infty} n|a_n|^2 r_1^{2n} \tag{1.7}$$

and

$$\sum_{n=-\infty}^{\infty} n|a_n|^2 r_2^{2n} \tag{1.8}$$

converge, as we now show for the case of the series (1.8). Writing

$$\sum_{n=-\infty}^{\infty} n|a_n|^2\rho^{2n} = \sum_{n=1}^{\infty} n|a_n|^2\rho^{2n} - \sum_{n=1}^{\infty} n|a_{-n}|^2\rho^{-2n} \qquad (r_1 < \rho < r_2),$$

and using (1.6), we see that

$$\sum_{n=1}^{\infty} n|a_n|^2\rho^{2n} < R_2^2 + \sum_{n=1}^{\infty} n|a_{-n}|^2\rho^{-2n} \qquad (r_1 < \rho < r_2).$$

Therefore

$$\sum_{n=1}^{N} n|a_n|^2\rho^{2n} < R_2^2 + \sum_{n=1}^{\infty} n|a_{-n}|^2\rho^{-2n} \qquad (r_1 < \rho < r_2)$$

for any integer $N > 0$, and hence, taking the limit as $\rho \to r_2-$, we find that

$$\sum_{n=1}^{N} n|a_n|^2 r_2^{2n} \leqslant R_2^2 + \sum_{n=1}^{\infty} n|a_{-n}|^2 r_2^{-2n},$$

which implies the convergence of (1.8), as well as the inequality

$$\sum_{n=-\infty}^{\infty} n|a_n|^2 r_2^{2n} \leqslant R_2^2. \tag{1.9}$$

In just the same way, we can prove the convergence of (1.7) and the inequality

$$\sum_{n=-\infty}^{\infty} n|a_n|^2 r_1^{2n} \geqslant R_1^2. \tag{1.10}$$

It follows from (1.9) and (1.10) that

$$\lambda^2 = \frac{\displaystyle\sum_{n=-\infty}^{\infty} n|a_n|^2 r_2^{2n}}{\displaystyle\sum_{n=-\infty}^{\infty} n|a_n|^2 r_1^{2n}} \leqslant \frac{R_2^2}{R_1^2}.$$

On the other hand, noting that

$$\lambda^2 - \frac{r_2^2}{r_1^2} = \frac{r_1^2 r_2^2 \sum_{n=-\infty}^{\infty} n|a_n|^2(r_2^{2n-2} - r_1^{2n-2})}{r_1^2 \sum_{n=-\infty}^{\infty} n|a_n|^2 r_1^{2n}} \geqslant 0,$$

since $n(r_2^{2n-2} - r_1^{2n-2}) \geqslant 0$ for all integral n, we conclude that

$$\frac{r_2^2}{r_1^2} \leqslant \lambda^2 \leqslant \frac{R_2^2}{R_1^2},$$

i.e.,

$$\frac{R_2}{R_1} \geqslant \frac{r_2}{r_1}. \tag{1.11}$$

However, the above argument obviously continues to hold if we interchange the role of the z and w-planes and of the domains G and D. Therefore

$$\frac{R_2}{R_1} \leqslant \frac{r_2}{r_1},$$

which, together with (1.11), implies the similarity condition (1.1). But then

$$\lambda^2 - \frac{r_2^2}{r_1^2} = \frac{r_1^2 r_2^2 \sum_{n=-\infty}^{\infty} n|a_n|^2(r_2^{2n-2} - r_1^{2n-2})}{r_1^2 \sum_{n=-\infty}^{\infty} n|a_n|^2 r_1^{2n}} = 0,$$

and hence $a_n = 0$ unless $n = 0$ or $n = 1$. Therefore

$$f(z) = a_0 + a_1 z,$$

i.e., $f(z)$ is a Möbius transformation, as asserted (in fact, an entire linear transformation).

To conclude the proof, we need only consider the case where $w = f(z)$ traverses Γ_ρ in the direction opposite to that in which the point z traverses γ_ρ. In this case, the transformation $w' = 1/w$ carries the annulus D into a new annulus

$$D' : \frac{1}{R_2} < |w'| < \frac{1}{R_1}$$

such that

$$w' = \frac{1}{w} = \frac{1}{f(z)} \tag{1.12}$$

is a conformal mapping of G onto D'. Let Γ'_ρ be the image of γ_ρ under the mapping (1.12). Then, since

$$\operatorname{Arg} w' = -\operatorname{Arg} f(z) = -\operatorname{Arg} w,$$

it follows that the point w' traverses Γ_ρ' in the direction opposite to that in which w traverses Γ_ρ, and hence in the same direction as z traverses γ_ρ. Therefore, by the same argument as before,

$$\frac{1/R_1}{1/R_2} = \frac{R_2}{R_1} = \frac{r_2}{r_1},$$

i.e., the annuli G and D are again similar. Moreover, we see that

$$\frac{1}{f(z)} = a_0 + a_1 z,$$

so that $f(z)$ is again a Möbius transformation. The proof of the theorem is now complete.

Remark. The domains G and D are homeomorphic, regardless of whether the condition (1.1) holds. In fact, the function

$$w = \left(R_2 \frac{|z| - r_1}{r_2 - r_1} + R_1 \frac{r_2 - |z|}{r_2 - r_1}\right) \frac{z}{|z|}$$

is a homeomorphism of G onto D. Nevertheless, G cannot be mapped conformally onto D unless the special condition (1.1) is satisfied.

2. Conformal Mapping of Simply Connected Domains

In view of Theorem 1.1, it is remarkable that any two *simply connected* domains in the extended plane, whose boundaries contain more than one point each, can always be mapped conformally onto each other.[5] To prove this, it is sufficient to show that every simply connected domain G can be mapped conformally onto some disk $K_R: |w| < R$ with its center at the origin. In fact, if $w = f(z)$ maps G conformally onto K_R, while $w = F(z)$ maps another domain D conformally onto the disk $K_{R'}: |w| < R'$, then the function

$$F^{-1}\left[\frac{R'}{R} f(z)\right]$$

maps G conformally onto D.

We now prove a fundamental theorem due to Riemann, on the existence of conformal mappings:

THEOREM 1.2 (*Riemann's mapping theorem*). *Every simply connected domain G in the extended plane whose boundary contains more than one point can be mapped conformally onto a disk with its center at the origin.*

[5] From now on, we shall confine ourselves to the study of this case, referring the reader interested in the theory of conformal mapping of multiply connected domains to the literature. See e.g., the books by Goluzin and Nehari (cited in the Bibliography on pp. 351–353). The work of M. V. Keldysh in this area is notable.

Proof. We begin by showing that the set of functions univalent on G contains bounded functions. (This set is nonempty, since any Möbius transformation is univalent on G.) First let G be bounded. Then the function $f(z) = z$ is obviously bounded. Next suppose G is unbounded, but has an exterior point z_0. Then the exterior of G contains some disk $|z - z_0| < \rho$, and the function

$$f(z) = \frac{1}{z - z_0}$$

is univalent on G and moreover bounded, since

$$|f(z)| < \frac{1}{\rho} \qquad (z \in G).$$

Finally suppose G is unbounded, but has no exterior. By hypothesis, G has at least two distinct boundary points α and β. Since G is simply connected, α and β belong to some continuum Γ which is the boundary of G. Consider the double-valued function

$$\varphi(z) = \sqrt{\frac{z - \alpha}{z - \beta}},$$

with two branch points α and β. Since G is obtained by excluding a continuum containing α and β from the extended plane, the function $\varphi(z)$ has two single-valued branches $\varphi_1(z)$ and $\varphi_2(z)$ on G, whose values at every point of G differ only in sign. The functions $\varphi_1(z)$ and $\varphi_2(z)$ are univalent on G, since it follows from

$$\sqrt{\frac{z_1 - \alpha}{z_1 - \beta}} = \sqrt{\frac{z_2 - \alpha}{z_2 - \beta}}$$

that

$$\frac{z_1 - \alpha}{z_1 - \beta} = \frac{z_2 - \alpha}{z_2 - \beta}$$

and hence

$$z_1 = z_2.$$

Therefore the functions $w = \varphi_1(z)$ and $w = \varphi_2(z)$ map G conformally onto two domains G_1 and G_2. Obviously $G_1 \cap G_2 = 0$, since, as just shown, the presence of a common point $w = \varphi_1(z_1) = \varphi_2(z_2)$ would imply $z_1 = z_2$, thereby contradicting the fact that $\varphi_1(z) = -\varphi_2(z)$ [note that $\varphi_1(z)$ and $\varphi_2(z)$ cannot equal 0 or ∞]. Therefore every point $w_0 \in G_2$ is an exterior point of G_1, i.e., there exists a disk $|w - w_0| < \rho$ contained in the exterior of G_1. But then the function

$$f(z) = \frac{1}{w - w_0} = \frac{1}{\varphi_1(z) - w_0}$$

is univalent and bounded on G, since $|f(z)| < 1/\rho$.

Thus, in every case, there exist bounded univalent functions on G. Let $f(z)$ be such a function, and let z_0 be a finite point of G. Then $f'(z_0) \neq 0$, since $f(z)$ is univalent on G (see Theorem II.4.1, Corollary 2), and hence

$$F(z) = \frac{f(z) - f(z_0)}{f'(z_0)}$$

is a bounded univalent function on G, which vanishes at z_0 and whose derivative at z_0 equals 1. Let E_{z_0} denote the set of all functions defined on G which have the same properties as $F(z)$. Then every $F(z) \in E_{z_0}$ has a finite least upper bound

$$M(F) = \sup_{z \in G} |F(z)| > 0.$$

Geometrically speaking, $M(F)$ is just the radius of the smallest disk with center at the origin which contains the image of G under the mapping $w = F(z)$. In general, we obtain different numbers $M(F)$ for different functions $F(z) \in E_{z_0}$. Our plan is to show that E_{z_0} contains a function $\psi(z)$ such that $M(\psi)$ equals the smallest value

$$R_{z_0} = \inf_{F \in E_{z_0}} M(F).$$

It will then turn out that the function $w = \psi(z)$ maps G conformally onto the disk $|w| < R_{z_0}$.

With this in mind, let $\{F_n(z)\}$ be a sequence of functions in E_{z_0} such that

$$\lim_{n \to \infty} M(F_n) = \inf_{F \in E_{z_0}} M(F) = R_{z_0}.$$

(Such a sequence exists by the very definition of the greatest lower bound.) This sequence is uniformly bounded on G, since the convergent numerical sequence $\{M(F_n)\}$ is bounded:

$$M(F_n) = \sup_{z \in G} |F_n(z)| \leqslant M \qquad (n = 1, 2, \ldots).$$

Therefore, by Montel's compactness criterion (Theorem I.17.17), the sequence $\{F_n(z)\}$ is compact on G, and hence contains a subsequence $\{F_{n_k}(z)\}$ which converges uniformly inside G to an analytic function $\psi(z)$.[6] For this function we have

$$\psi(z_0) = \lim_{k \to \infty} F_{n_k}(z_0) = 0, \qquad \psi'(z_0) = \lim_{k \to \infty} F'_{n_k}(z_0) = 1,$$

which implies that $\psi(z) \not\equiv \text{const}$. It follows from Theorem II.4.4 that $\psi(z)$ is univalent on G. Moreover

$$\psi(z) = \lim_{k \to \infty} F_{n_k}(z), \qquad |F_{n_k}(z)| \leqslant M(F_{n_k}) \to R_{z_0},$$

[6] Here and henceforth "inside G" means "on every compact subset of G," as in the definition on p. I.411.

and hence, given any $\varepsilon > 0$,

$$|\psi(z)| < |F_{n_k}(z)| + \frac{\varepsilon}{2} < R_{z_0} + \varepsilon$$

for all sufficiently large n_k. Therefore

$$|\psi(z)| \leqslant R_{z_0} \qquad (z \in G), \tag{1.13}$$

since $\varepsilon > 0$ is arbitrary. On the other hand $\psi(z) \in E_{z_0}$, since $\psi(z)$ is bounded and univalent on G, and

$$\psi(z_0) = 0, \qquad \psi'(z_0) = 1.$$

It follows that

$$\sup_{z \in G} |\psi(z)| \geqslant R_{z_0},$$

which, together with (1.13), implies

$$\sup_{z \in G} |\psi(z)| = R_{z_0}.$$

The rest of the proof consists in showing that the function $w = \psi(z)$ maps the domain G conformally onto the disk $K_0: |w| < R_{z_0}$. Since we already know that $\psi(z)$ is univalent on G and that $\psi(G)$ is contained in K_0, we need only prove that every point of the disk K_0 belongs to $\psi(G)$. Suppose the opposite. Then K_0 must contain boundary points of $\psi(G)$. Let w_0 $(0 < |w_0| < R_{z_0})$ be such a point, and form the following functions, each involving the preceding one:

$$w = \psi(z), \quad w_1 = R_{z_0}^2 \frac{w - w_0}{R_{z_0}^2 - \bar{w}_0 w} = \psi_1(z), \quad w_2 = \sqrt{R_{z_0} w_1} = \psi_2(z),$$

$$w_3 = R_{z_0}^2 \frac{w_2 - \psi_2(z_0)}{R_{z_0}^2 - \overline{\psi_2(z_0)} w_2} = \psi_3(z), \quad w_4 = \frac{w_3}{\psi_3'(z_0)} = \psi_4(z)$$

(w_2 involves a fixed branch of $\sqrt{w_1}$). Each of the functions $w_j = \psi_j(z)$, regarded as a function of w_{j-1} $(w_0 = w)$, is univalent on the domain $\psi_{j-1}(G)$ $[\psi_0(G) = \psi(G)]$. Therefore all the functions, regarded as functions of z, are univalent on the domain G. The function $w_1(w)$ transforms the disk K_0 into itself, carrying the point w_0 into the origin $w_1 = 0$ and the domain $\psi(G)$ into a domain $\psi_1(G)$ contained in K_0, with $w_1 = 0$ as a boundary point.[7] The function $w_2(w_1)$ has two single-valued analytic branches on the domain $\psi_1(G)$. Choosing one of these branches, we note that it takes all its values in the disk K_0 and carries the point $w_1 = 0$ into the point $w_2 = 0$. Thus $\psi_2(G) \subset K_0$, and $w_2 = 0$ is a boundary point of $\psi_2(G)$. The function $w_3(w_2)$ transforms K_0 into itself and carries $\psi_2(z_0)$ into the origin. Therefore the univalent function $\psi_3(z)$

[7] The same symbol K_0 is used for all the disks $|w| < R_{z_0}$, $|w_1| < R_{z_0}, \ldots, |w_4| < R_{z_0}$.

vanishes at the point z_0 and maps G onto a domain $\psi_3(G) \subset K_0$. Calculating the derivative $\psi_3'(z_0)$, we obtain

$$\psi_3'(z_0) = \left.\frac{dw}{dz}\right|_{z=z_0} \cdot \left.\frac{dw_1}{dw}\right|_{w=\psi(z_0)=0} \cdot \left.\frac{dw_2}{dw_1}\right|_{w_1=\psi_1(z_0)=-w_0} \cdot \left.\frac{dw_3}{dw_2}\right|_{w_2=\psi_2(z_0)=\sqrt{-R_{z_0}w_0}}$$

$$= 1 \cdot \frac{R_{z_0}^2 - |w_0|^2}{R_{z_0}^2} \cdot \frac{\sqrt{R_{z_0}}}{2\sqrt{-w_0}} \cdot \frac{R_{z_0}^2}{R_{z_0}^2 - R_{z_0}|w_0|} = \frac{R_{z_0} + |w_0|}{2\sqrt{-R_{z_0}w_0}}.$$

It follows that $|\psi_3'(z_0)| > 1$, since $|w_0| < R_{z_0}$, and hence $\psi_3(z) \notin E_{z_0}$. However, dividing $\psi_3(z)$ by $\psi_3'(z_0)$, we obtain a function $\psi_4(z) \in E_{z_0}$ [$\psi_4(z)$ is univalent on G, $\psi_4(z_0) = 0$, $\psi_4'(z_0) = 1$]. Clearly, $\psi_4(z)$ maps G onto a domain $\psi_4(G)$ contained in the disk

$$|w_4| < \frac{R_{z_0}}{|\psi_3'(z_0)|} < R_{z_0},$$

and hence

$$\sup_{z \in G} |\psi_4(z)| \leqslant \frac{R_{z_0}}{|\psi_3'(z_0)|} < R_{z_0}.$$

But this contradicts the definition of the number R_{z_0}, as the greatest lower bound of the numbers

$$\sup_{z \in G} |F(z)|, \qquad F(z) \in E_{z_0},$$

thereby proving the theorem.

Remark. In the course of the proof, we found that the function $w = \psi(z)$ accomplishing the conformal mapping of G onto K_0 can be made to satisfy the extra conditions

$$\psi(z_0) = 0, \qquad \psi'(z_0) = 1,$$

where z_0 is an arbitrary finite point of G. The disk K_0 found in the proof is centered at the origin and has a well-defined radius R_{z_0}, called the *conformal radius* of the domain G (relative to the point z_0). Replacing $\psi(z)$ by the function

$$f(z) = \frac{1}{R_{z_0}} \psi(z),$$

we obtain a conformal mapping of G onto the unit disk $|w| < 1$, where the function $f(z)$ accomplishing this mapping satisfies the conditions

$$f(z_0) = 0, \qquad f'(z_0) > 0.$$

Geometrically, this means that the point $z_0 \in G$ goes into the center of the unit disk, while the tangents to all curves passing through the point z_0 experience no rotation when subjected to the transformation $w = f(z)$ mapping them into curves passing through the center of the disk.

THEOREM 1.3 (*Uniqueness theorem for conformal mapping*).[8] *Let G be the same as in Theorem* 1.2, *and let z_0 be an arbitrary finite point of G. Then there exists a unique function $w = f(z)$ which maps G conformally onto the disk $|w| < 1$ and satisfies the conditions*

$$f(z_0) = 0, \qquad f'(z_0) > 0.$$

Proof. Suppose $w = g(z)$ is another function satisfying the same conditions as $w = f(z)$. Then the function

$$\varphi(w) = f[g^{-1}(w)]$$

is analytic on the unit disk, satisfies the conditions

$$\varphi(0) = f[g^{-1}(0)] = f(z_0) = 0, \qquad \varphi'(0) = \frac{1}{g'(z_0)} f'(z_0) > 0,$$

and maps the unit disk onto itself conformally. Therefore, according to Schwarz's lemma (Theorem I.17.8),

$$|\varphi(w)| \leqslant |w|,$$

i.e.,

$$|f(z)| \leqslant |g(z)| \tag{1.14}$$

for all $z \in G$. But then, interchanging the roles of $f(z)$ and $g(z)$, we obtain

$$|g(z)| \leqslant |f(z)|,$$

which together with (1.14) implies

$$|f(z)| = |g(z)|,$$

or equivalently

$$|\varphi(w)| = |w|. \tag{1.15}$$

Applying Schwarz's lemma again to (1.15), we conclude that

$$\varphi(w) = e^{i\alpha} w.$$

But $\varphi'(0) > 0$ and hence $e^{i\alpha} = 1$, i.e.,

$$\varphi(w) = w.$$

It follows that

$$f(z) = g(z)$$

for all $z \in G$, as asserted.

---

[8] Theorem 1.3 was anticipated in Problem II.5.22. Clearly, there are infinitely many conformal mappings of G onto $|w| < 1$ unless we impose further conditions on the mappings. This follows from the fact that there are infinitely many conformal mappings of the unit disk onto itself (accomplished by Möbius transformations).

3. Basic Properties of Univalent Functions

In this section, we pursue the study of univalent functions begun in Chap. II.4.

THEOREM 1.4 ("*Area theorem*"). *If the function*

$$f(z) = z + \alpha_0 + \alpha_1 z^{-1} + \cdots + \alpha_n z^{-n} + \cdots$$

is univalent outside the unit circle $|z| = 1$, *then*

$$\sum_{n=1}^{\infty} n|\alpha_n|^2 \leqslant 1. \tag{1.16}$$

Proof. The function $w = f(z)$ maps the circle $\gamma_\rho: |z| = \rho > 1$ onto a closed Jordan curve Γ_ρ, bounding a domain of area S_ρ, and one circuit around γ_ρ in the positive direction corresponds to one circuit around Γ_ρ in the positive direction (why?). As on p. 5, the area S_ρ is given by the expression

$$\begin{aligned} S_\rho &= \frac{1}{2}\int_{\Gamma_\rho} (u\,dv - v\,du) = \frac{1}{2}\int_0^{2\pi} \operatorname{Im}\left\{\overline{f(\rho e^{i\theta})}\,\frac{\partial f(\rho e^{i\theta})}{\partial \theta}\right\} d\theta \\ &= \frac{1}{2}\int_0^{2\pi} \operatorname{Im}\{(\rho e^{-i\theta} + \bar{\alpha}_0 + \bar{\alpha}_1\rho^{-1}e^{i\theta} + \bar{\alpha}_2\rho^{-2}e^{2i\theta} + \cdots) \\ &\qquad \times i(\rho e^{i\theta} - \alpha_1\rho^{-1}e^{-i\theta} - 2\alpha_2\rho^{-2}e^{-2i\theta} - \cdots)\}\,d\theta \\ &= \operatorname{Im}\left\{\frac{2\pi i}{2}(\rho^2 - |\alpha_1|^2\rho^{-2} - 2|\alpha_2|^2\rho^{-4} - \cdots - n|\alpha_n|^2\rho^{-2n} - \cdots)\right\} \\ &= \pi\left(\rho^2 - \sum_{n=1}^{\infty} n|\alpha_n|^2\rho^{-2n}\right) > 0. \end{aligned}$$

It follows that

$$\sum_{n=1}^{\infty} n|\alpha_n|^2\rho^{-2n} < \rho^2$$

for any $\rho > 1$. Therefore

$$\sum_{n=1}^{N} n|\alpha_n|^2\rho^{-2n} < \rho^2$$

for any integer $N > 0$, and hence, taking the limit as $\rho \to 1+$, we obtain first

$$\sum_{n=1}^{N} n|\alpha_n|^2 \leqslant 1,$$

and then (1.16) after passing to the limit $N \to \infty$.

THEOREM 1.5. *Let the function*

$$f(z) = z + a_2 z^2 + \cdots + a_n z^n + \cdots$$

be univalent on the unit disk $|z| < 1$. *Then*

$$|a_2| \leqslant 2, \tag{1.17}$$

where the equality holds only for the function

$$f(z) = \frac{z}{(1 - e^{i\alpha}z)^2} = z + 2e^{i\alpha}z^2 + 3e^{2i\alpha}z^3 + \cdots + ne^{(n-1)i\alpha}z^n + \cdots, \tag{1.18}$$

which maps the unit disk conformally onto the domain G whose boundary is the ray of inclinaton $\pi - \alpha$ *obtained by joining the point* $-\frac{1}{4}e^{-i\alpha}$ *to the point at infinity.*

Proof. Consider the function

$$\varphi(z) = 1 + a_2z^2 + a_3z^4 + \cdots + a_nz^{2n-2} + \cdots.$$

This function is single-valued, nonvanishing and analytic on the unit disk. In fact $\varphi(0) = 1 \neq 0$, while if $z \neq 0$, then

$$\varphi(z) = \frac{f(z^2)}{z^2},$$

and since $f(z^2) \neq 0$ if $z \neq 0$ (by the univalence), we must have $\varphi(z) \neq 0$ if $z \neq 0$. Therefore the double-valued function

$$\psi(z) = \sqrt{\frac{f(z^2)}{z^2}}$$

has two single-valued analytic branches on the unit disk. Choosing the branch which takes the value 1 for $z = 0$, we have

$$\begin{aligned}\psi(z) &= [1 + (a_2z^2 + \cdots)]^{1/2} \\ &= 1 + \tfrac{1}{2}(a_2z^2 + \cdots) - \tfrac{1}{8}(a_2z^2 + \cdots)^2 + \cdots \\ &= 1 + \frac{a_2}{2}z^2 + \cdots\end{aligned}$$

[this expansion converges on the unit disk, by the analyticity of $\psi(z)$]. Obviously, the function $\psi(z)$ is even.

We now form the odd function

$$\chi(z) = z + \frac{a_2}{2}z^3 + \cdots = z\psi(z) = z\sqrt{\frac{f(z^2)}{z^2}} = \sqrt{f(z^2)}. \tag{1.19}$$

This function is univalent on the unit disk, since if $\chi(z_1) = \chi(z_2)$, then $f(z_1^2) = f(z_2^2)$, i.e., $z_1 = \pm z_2$. But if $z_1 = -z_2 \neq 0$, then $\chi(z_1) = -\chi(z_2)$, and hence the equality $\chi(z_1) = \chi(z_2)$ is impossible [the function $\chi(z)$ vanishes with $f(z^2)$, i.e., only for $z = 0$]. Thus we must have $z_1 = z_2$, which implies that $\chi(z)$ is univalent, as asserted.

Next we replace z by $1/\zeta$ and

$$\chi(z) = \chi\left(\frac{1}{\zeta}\right)$$

by

$$F(\zeta) = \frac{1}{\chi\left(\frac{1}{\zeta}\right)}.$$

Since, in addition to $\chi(z)$, this involves only Möbius transformations, which are univalent, the function $F(\zeta)$ must be univalent on the exterior of the unit circle $|\zeta| = 1$ (if $|z| < 1$, then $|\zeta| > 1$). For $F(\zeta)$ we have the expansion

$$F(\zeta) = \frac{1}{\frac{1}{\zeta} + \frac{a_2}{2}\frac{1}{\zeta^3} + \cdots} = \frac{\zeta}{1 + \frac{a_2}{2}\frac{1}{\zeta^2} + \cdots}$$

$$= \zeta\left(1 - \frac{a_2}{2}\frac{1}{\zeta^2} + \cdots\right) = \zeta - \frac{a_2}{2}\zeta^{-1} + \cdots.$$

Therefore, according to Theorem 1.4,

$$\left|\frac{a_2}{2}\right|^2 + \cdots \leqslant 1,$$

and hence $|a_2| \leqslant 2$, which agrees with (1.17).

Finally suppose $|a_2| = 2$, so that $a_2 = 2e^{i\alpha}$, where α is real. Then all the coefficients in the expansion of $F(\zeta)$, starting from that of ζ^{-2}, must vanish. Therefore

$$F(\zeta) = \zeta - e^{i\alpha}\zeta^{-1},$$

which implies

$$\chi(z) = \chi\left(\frac{1}{\zeta}\right) = \frac{1}{F(\zeta)} = \frac{1}{\zeta - e^{i\alpha}\zeta^{-1}} = \frac{z}{1 - e^{i\alpha}z^2},$$

or

$$f(z^2) = [\chi(z)]^2 = \frac{z^2}{(1 - e^{i\alpha}z^2)^2},$$

and hence

$$f(z) = \frac{z}{(1 - e^{i\alpha}z)^2},$$

which agrees with (1.18), where the indicated Taylor series expansion is easily verified. To see that the function $f(z)$ maps the unit disk onto the domain G figuring in the statement of the theorem, we need only carry out the following consecutive transformations:

$$z_1 = -e^{i\alpha}z, \qquad z_2 = \frac{1}{2}\left(z_1 + \frac{1}{z_1}\right), \qquad z_3 = -2e^{i\alpha}(z_2 + 1),$$

$$w = \frac{1}{z_3} = f(z).$$

The proof is now complete.

THEOREM 1.6. *Let G be the image of the unit disk $|z| < 1$ under the univalent mapping*

$$w = f(z) = z + a_2z^2 + \cdots + a_nz^n + \cdots.$$

Then G always contains the disk $|w| < \frac{1}{4}$.

Proof. Let w_0 be an arbitrary point not in G. Then the function

$$\varphi(z) = \frac{f(z)}{1 - \frac{1}{w_0}f(z)},$$

which like $f(z)$ is univalent on the unit disk, has the expansion

$$\varphi(z) = \frac{z + a_2z^2 + \cdots}{1 - \frac{1}{w_0}z - \frac{a_2}{w_0}z^2 - \cdots} = z + \left(a_2 + \frac{1}{w_0}\right)z^2 + \cdots.$$

According to Theorem 1.5, we must have

$$\left|a_2 + \frac{1}{w_0}\right| \leqslant 2,$$

and hence

$$\left|\frac{1}{w_0}\right| \leqslant 2 + |a_2| \leqslant 4,$$

i.e.,

$$|w_0| \geqslant \tfrac{1}{4}.$$

Therefore the disk $|w| < \frac{1}{4}$ is contained in G, as asserted.

Remark. The constant $\frac{1}{4}$ cannot be made larger, as shown by choosing $f(z)$ to be the function (1.18).

THEOREM 1.7 ("*Distortion theorem*"). *If the function*

$$f(z) = z + a_2z^2 + \cdots + a_nz^n + \cdots$$

is univalent on the unit disk $|z| < 1$, then the magnification $|f'(z)|$ (the "distortion") under the mapping $w = f(z)$ satisfies the inequality

$$\frac{1 - |z|}{(1 + |z|)^3} \leqslant |f'(z)| \leqslant \frac{1 + |z|}{(1 - |z|)^3} \qquad (|z| < 1), \tag{1.20}$$

where neither equality can hold for $z \neq 0$ unless

$$f(z) = \frac{z}{(1 - e^{i\alpha}z)^2}.$$

Proof. Choosing an arbitrary point $z = re^{i\theta}$ of modulus less than 1, we consider the mapping

$$\zeta' = \frac{\zeta - z}{\zeta\bar{z} - 1}, \tag{1.21}$$

which carries the point z into the original and the unit disk $|\zeta| < 1$ into itself. Then

$$\zeta = \frac{\zeta' - z}{\zeta'\bar{z} - 1},$$

and hence, under the mapping (1.21), $f(\zeta)$ goes into the function

$$\varphi(\zeta') = f\left(\frac{\zeta' - z}{\zeta'\bar{z} - 1}\right),$$

which is univalent on the unit disk $|\zeta'| < 1$. Clearly

$$\varphi(0) = f(z), \qquad \varphi'(0) = -f'(z)(1 - |z|^2),$$
$$\varphi''(0) = f''(z)(1 - |z|^2)^2 - 2\bar{z}f'(z)(1 - |z|^2), \ldots,$$

and hence

$$\varphi(\zeta') = f(z) - f'(z)(1 - |z|^2)\zeta' + \tfrac{1}{2}[f''(z)(1 - |z|^2)^2 - 2\bar{z}f'(z)(1 - |z|^2)]\zeta'^2 + \cdots$$

or

$$\psi(\zeta') \equiv \frac{\varphi(\zeta') - f(z)}{-f'(z)(1 - |z|^2)} = \zeta' - \frac{1}{2}\left[\frac{f''(z)}{f'(z)}(1 - |z|^2) - 2\bar{z}\right]\zeta'^2 + \cdots.$$

Applying Theorem 1.5 to the univalent function $\psi(\zeta')$, we obtain

$$\frac{1}{2}\left|\frac{f''(z)}{f'(z)}(1 - |z|^2) - 2\bar{z}\right| \leqslant 2, \tag{1.22}$$

or

$$\left|z\frac{f''(z)}{f'(z)} - \frac{2r^2}{1 - r^2}\right| \leqslant \frac{4r}{1 - r^2}$$

($r = |z|$), and therefore

$$-\frac{4r}{1 - r^2} \leqslant \operatorname{Re}\left\{z\frac{f''(z)}{f'(z)} - \frac{2r^2}{1 - r^2}\right\} \leqslant \frac{4r}{1 - r^2},$$
$$-\frac{4r}{1 - r^2} \leqslant \operatorname{Im}\left\{z\frac{f''(z)}{f'(z)} - \frac{2r^2}{1 - r^2}\right\} \leqslant \frac{4r}{1 - r^2}.$$

Since

$$z\frac{f''(z)}{f'(z)} = z\frac{d\operatorname{Ln} f'(z)}{dz} = r\left[\frac{\partial \ln |f'(z)|}{\partial r} + i\frac{\partial \operatorname{Arg} f'(z)}{\partial r}\right] \tag{1.23}$$

[cf. (I.7.24)], we find that

$$-\frac{4r}{1 - r^2} \leqslant r\frac{\partial \ln |f'(z)|}{\partial r} - \frac{2r^2}{1 - r^2} \leqslant \frac{4r}{1 - r^2},$$
$$-\frac{4r}{1 - r^2} \leqslant r\frac{\partial \operatorname{Arg} f'(z)}{\partial r} \leqslant \frac{4r}{1 - r^2},$$

i.e.,

$$\frac{2r-4}{1-r^2} \leqslant \frac{\partial \ln |f'(z)|}{\partial r} \leqslant \frac{2r+4}{1-r^2}, \tag{1.24}$$

$$-\frac{4}{1-r^2} \leqslant \frac{\partial \operatorname{Arg} f'(z)}{\partial r} \leqslant \frac{4}{1-r^2}. \tag{1.25}$$

Integration of (1.24) with respect to r from 0 to $r = |z|$ gives

$$\ln \frac{1-r}{(1+r)^3} \leqslant \ln |f'(z)| \leqslant \ln \frac{1+r}{(1-r)^3},$$

which is just another way of writing (1.20).

Finally, we observe that equality in the left or right-hand side of (1.20) for any $z \neq 0$ implies equality in the corresponding side of (1.24) for all points of the line segment joining 0 and z, and similarly in the corresponding sides of (1.22) [written as two inequalities]. Setting $z = 0$ in (1.22) and noting that $f'(0) = 1$, we obtain

$$|a_2| = \tfrac{1}{2}|f''(0)| = 2.$$

It follows from Theorem 1.5 that

$$f(z) = \frac{z}{(1-e^{i\alpha}z)^2}, \tag{1.26}$$

as asserted. The fact that the equalities in (1.20) are actually achieved for this function is easily verified. Indeed, since

$$f'(z) = \frac{1+e^{i\alpha}z}{(1-e^{i\alpha}z)^3},$$

we have

$$f'(z) = \frac{1+|z|}{(1-|z|)^3}$$

for $z = |z|e^{-i\alpha}$, and

$$f'(z) = \frac{1-|z|}{(1+|z|)^3}$$

for $z = -|z|e^{-i\alpha}$.

COROLLARY. *If $f(z)$ is the same as in Theorem* 1.7, *then*

$$|\arg f'(z)| \leqslant 2 \ln \frac{1+|z|}{1-|z|} \qquad (|z| < 1). \tag{1.27}$$

Proof. Integrate (1.25) from 0 to $r = |z|$, noting that $f'(0) = 1$ and choosing the single-valued branch of $\operatorname{Arg} f'(z)$ which vanishes for $z = 0$.

Remark. Unlike (1.20), the inequality (1.27) is not the best possible estimate. The definitive estimate of $\arg f'(z)$, obtained by another, more complicated method, is

$$|\arg f'(z)| \leqslant \begin{cases} 4 \arcsin |z| & \text{if } |z| \leqslant \dfrac{1}{\sqrt{2}}, \\ \pi + \ln \dfrac{|z|^2}{1 - |z|^2} & \text{if } \dfrac{1}{\sqrt{2}} \leqslant |z| < 1. \end{cases} \tag{1.28}$$

This result, known as the *rotation theorem*, is due to Goluzin.[9] The estimates (1.28) are *sharp* in the sense that there exist univalent functions, different from (1.26), for which $|\arg f'(z)|$ takes values arbitrarily close to the right-hand side of (1.28) at any preassigned point of the unit disk.

Next we use the first inequality in (1.24) to deduce a simple geometric property of conformal mappings:

THEOREM 1.8. *If the function*

$$f(z) = z + a_2 z^2 + \cdots + a_n z^n + \cdots$$

is univalent on the unit disk $|z| < 1$, *then the image of every circle*

$$|z| = r \leqslant 2 - \sqrt{3}$$

is a convex curve.[10]

Proof. Since

$$r \frac{\partial \ln |f'(z)|}{\partial r} = \frac{\partial \operatorname{Arg} f'(z)}{\partial \theta} = \operatorname{Re}\left\{ z \frac{d \operatorname{Ln} f'(z)}{dz} \right\} = \operatorname{Re}\left\{ z \frac{f''(z)}{f'(z)} \right\}$$

[cf. (1.23) and (I.7.22)], it follows from (1.24) that

$$\frac{\partial \operatorname{Arg} f'(z)}{\partial \theta} \geqslant \frac{2r^2 - 4r}{1 - r^2}. \tag{1.29}$$

Let $\Gamma: w = f(re^{i\theta})$, $0 \leqslant \theta \leqslant 2\pi$ be the image of the circle $\gamma: |z| = r < 1$ under the mapping $w = f(z)$. Since the tangent to γ at the point $z = re^{i\theta}$ has inclination

$$\theta + \frac{\pi}{2},$$

the tangent to Γ at the corresponding point $f(re^{i\theta})$ has inclination

$$\varphi = \theta + \frac{\pi}{2} + \operatorname{Arg} f'(re^{i\theta}).$$

---

[9] See G. M. Goluzin, *op. cit.*, Chap. 4, Sec. 1.

[10] A closed Jordan curve is said to be *convex* if every straight line intersecting both the interior and the exterior of the curve intersects the curve itself in just two points.

The sign of the derivative

$$\frac{\partial \varphi}{\partial \theta} = 1 + \frac{\partial \operatorname{Arg} f'(re^{i\theta})}{\partial \theta} \tag{1.30}$$

determines how the inclination of the tangent to Γ at the variable point $w = f(re^{i\theta})$ changes with w (note that w describes Γ in the positive direction as θ increases). In particular, if $\partial\varphi/\partial\theta \geqslant 0$ in some interval (θ', θ''), then the corresponding arc of Γ is concave towards the point $w = 0$ (convex when seen from outside Γ), while if $\partial\varphi/\partial\theta \leqslant 0$ in (θ', θ''), this arc is convex towards $w = 0$ (concave from outside Γ). But information about the sign of the derivative (1.30) can be deduced from (1.29). In fact

$$\frac{\partial \varphi}{\partial \theta} \geqslant \frac{r^2 - 4r + 1}{1 - r^2} = \frac{(r - 2 + \sqrt{3})(r - 2 - \sqrt{3})}{1 - r^2},$$

and hence $\partial\varphi/\partial\theta \geqslant 0$ for all $\theta \in [0, 2\pi]$ if

$$r \leqslant 2 - \sqrt{3} = 0.2679\ldots$$

This completes the proof, since then Γ is convex.

Remark. For the function (1.26) we have

$$z\frac{f''(z)}{f'(z)} = \frac{2(e^{i\alpha}z)^2 + 4e^{i\alpha}z}{1 - (e^{i\alpha}z)^2},$$

which for $z = -re^{-i\alpha}$ takes real values

$$\frac{2r^2 - 4r}{1 - r^2}.$$

Therefore, in this case

$$\frac{\partial \varphi}{\partial \theta} = 1 + \frac{\partial \operatorname{Arg} f'(z)}{\partial \theta} = 1 + \operatorname{Re}\left\{z\frac{f''(z)}{f'(z)}\right\}$$
$$= \frac{r^2 - 4r + 1}{1 - r^2} = \frac{(r - 2 + \sqrt{3})(r - 2 - \sqrt{3})}{1 - r^2},$$

from which it is clear that $\partial\varphi/\partial\theta < 0$ if $r > 2 - \sqrt{3}$ $(r < 1)$, i.e., the images of the circles $|z| = r > 2 - \sqrt{3}$ $(r < 1)$ under the mapping (1.26) are no longer convex.

THEOREM 1.9. *If the function*

$$f(z) = z + a_2z^2 + \cdots + a_nz^n + \cdots$$

is univalent on the unit disk $|z| < 1$, *then*

$$\frac{|z|}{(1 + |z|)^2} \leqslant |f(z)| \leqslant \frac{|z|}{(1 - |z|)^2} \qquad (|z| < 1), \tag{1.31}$$

where neither equality can hold for $z \neq 0$ *unless* $f(z)$ *is given by* (1.26).

Proof. According to Theorem 1.7,

$$|f'(z)| \leqslant \frac{1 + |z|}{(1 - |z|)^3}.$$

Integrating both sides of this inequality along the radius joining the points 0 and z, we find that

$$|f(z)| = \left|\int_0^z f'(z)\, dz\right| = \int_0^{|z|} |f'(z)|\, d|z|$$
$$\leqslant \int_0^{|z|} \frac{1 + |z|}{(1 - |z|)^3}\, d|z| = \frac{|z|}{(1 - |z|)^2}$$

(cf. Problem I.12.20), which is the second of the inequalities (1.31). To obtain the first inequality, consider the image Γ of the circle $|z| = r$ under the mapping $w = f(z)$. Clearly Γ is a closed Jordan curve in the w-plane, whose interior contains the point $w = 0$. Let R be the distance between $w = 0$ and Γ. Then Γ must contain at least one point $w_1 = f(z_1)$ (where $|z_1| = r$) such that $|f(z_1)| = R$. The line segment Δ joining the points 0 and $f(z_1)$ is contained in the image of the unit disk under the mapping $w = f(z)$, and hence the inverse image δ of this segment is contained in the unit disk. Thus, for $|z| = r$, we use the first of the inequalities (1.20) to obtain

$$|f(z)| \geqslant R = |f(z_1)| = \int_\Delta |dw| = \int_\delta |f'(z)|\, |dz| = \int_\delta |f'(z)| \left|\frac{dz}{d|z|}\right| d|z|$$
$$\geqslant \int_0^{|z|} |f'(z)|\, d|z| \geqslant \int_0^{|z|} \frac{1 - |z|}{(1 + |z|)^3}\, d|z| = \frac{|z|}{(1 + |z|)^2},$$

as required. Finally, it is obvious that equality can hold in (1.31) only if equality holds in (1.20), and then, as we know, $f(z)$ is given by (1.26).

THEOREM 1.10. *Let G be the domain onto which the univalent function*

$$f(z) = z + a_2 z^2 + \cdots + a_n z^n + \cdots$$

maps the unit disk $|z| < 1$. Then the (inner) area of G equals

$$S = \pi[1 + 2|a_2|^2 + \cdots + n|a_n|^2 + \cdots] \geqslant \pi, \tag{1.32}$$

where $S = \pi$ if and only if $f(z) = z$.

Proof. Let Γ_r be the image of the circle $|z| = r < 1$. Then Γ_r is a closed rectifiable Jordan curve whose interior G_r has a well-defined area S_r, given by the formula

$$S_r = \iint_{|z| \leqslant r} |f'(z)|^2 r\, dr\, d\theta \tag{1.33}$$

(see Problem 1.8). Moreover, it is clear that $\overline{G}_r \subset G_{r'} \subset \overline{G}_{r'} \subset G$ and $S_r < S_{r'}$ if $r < r'$. We now show that the domains G_r approximate G

from the inside in the sense that any given compact set $E \subset G$ is contained in G_r for all sufficiently large $r > r(E)$. Suppose this assertion is false. Then there exists an increasing sequence of positive numbers $\{r_n\}$ converging to 1 and a corresponding sequence of points $\{w_n\}$ in E such that $w_n \notin G_{r_n}$ $(n = 1, 2, \ldots)$. The sequence $\{w_n\}$ has a limit point $w_0 \in E \subset G$. The inverse image $z_0 = f^{-1}(w_0)$ belongs to the unit disk, and hence some neighborhood $\mathcal{N}(z_0)$ is contained in every disk $|z| < r_n$ with sufficiently large $n > N$. It follows that $f[\mathcal{N}(z_0)] \subset G_{r_n}$ for all $n > N$. Since $f[\mathcal{N}(z_0)]$ is a domain, there is a neighborhood $\mathcal{N}(w_0)$ such that $\mathcal{N}(w_0) \subset f[\mathcal{N}(z_0)]$. But $\mathcal{N}(w_0)$ contains points w_n with arbitrarily large values of n, and hence every point $w_n \in \mathcal{N}(w_0)$ with $n > N$ belongs to G_{r_n}, contrary to the definition of the sequence $\{w_n\}$. This contradiction shows that E is contained in some domain G_{r_ν} and hence in all domains G_r with $r > r_\nu$.

Now let $\mathcal{Q}$ be any family of closed squares constructed in the w-plane by drawing lines parallel to the coordinate axes, and let F be the union of all the squares of $\mathcal{Q}$ contained in G. By the *inner area* of G, we mean the quantity

$$S = \sup_{\mathcal{Q}} \text{Area of } F, \tag{1.34}$$

where the least upper bound is taken over all possible families $\mathcal{Q}$ (S may be infinite). Since F is a compact subset of G, we can always find a number $r_0 < 1$ such that $F \subset G_r$ for $r > r_0$. On the other hand, given any G_r, we can always find a set F of the type just described such that $G_r \subset F \subset G$ (why?). Together with the definition (1.34), this implies

$$S = \lim_{r \to 1-} S_r.$$

Returning to (1.33) and noting that $f'(z)$ has the series expansion

$$f'(z) = f'(re^{i\theta}) = 1 + \sum_{n=2}^{\infty} n a_n r^{n-1} e^{i(n-1)\theta},$$

we find that

$$\begin{aligned} S_r &= \int_0^r \int_0^{2\pi} \left(1 + \sum_{n=2}^{\infty} n a_n r^{n-1} e^{i(n-1)\theta}\right)\left(1 + \sum_{n=2}^{\infty} n \bar{a}_n r^{n-1} e^{-i(n-1)\theta}\right) r \, dr \, d\theta \\ &= \int_0^r 2\pi\left(r + \sum_{n=2}^{\infty} n^2 |a_n|^2 r^{2n-1}\right) dr = \pi\left(r^2 + \sum_{n=2}^{\infty} n |a_n|^2 r^{2n}\right). \end{aligned}$$

Suppose we replace r by 1 in the last series on the right, thereby obtaining

$$\pi\left(1 + \sum_{n=2}^{\infty} n |a_n|^2\right). \tag{1.35}$$

Then if (1.35) converges, it follows from Theorems I.15.2 and I.15.3 that S_r is a continuous function of r in [0, 1], and hence

$$S = \lim_{r \to 1-} S_r = \pi\left(1 + \sum_{n=2}^{\infty} n|a_n|^2\right).$$

On the other hand, if (1.35) diverges, then given any $A > 0$, however large, we can find an integer $N_0(A) > 0$ such that

$$\pi\left(1 + \sum_{n=2}^{N} n|a_n|^2\right) > A$$

for all $N > N_0(A)$. Therefore

$$\pi\left(1 + \sum_{n=2}^{N} n|a_n|^2 r^{2n}\right) > A$$

for all $N > N_0(A)$ and r sufficiently close to 1. But this implies

$$S_r = \pi\left(1 + \sum_{n=2}^{\infty} n|a_n|^2 r^{2n}\right) > A$$

for all r sufficiently close to 1, and hence

$$S = \lim_{r \to 1-} S_r = \infty.$$

Thus, whether finite or not, the area of the domain G is given by the formula (1.32), as asserted. Moreover, it is obvious that $S = \pi$ if and only if

$$a_2 = \cdots = a_n = \cdots = 0, \text{ i.e., } f(z) = z.$$

This completes the proof.

Remark. Thus a conformal mapping of the disk $|z| < 1$ by a function $f(z)$ such that $f(0) = 0, f'(0) = 1$ always leads to a domain of area greater than that of the disk, except when $f(z)$ reduces to the identity transformation. Naturally, this result can be extended to conformal mappings of a disk whose radius and center are arbitrary, provided only that the derivative of the function accomplishing the mapping equals 1 at the center of the disk.

THEOREM 1.11. *If the function*

$$f(z) = z + a_2 z^2 + \cdots + a_n z^n + \cdots$$

is univalent on the unit disk $|z| < 1$, *then*

$$\frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|\, d\theta \leqslant \frac{r}{1 - r} \qquad (r < 1). \tag{1.36}$$

Proof. Consider the odd univalent function

$$\chi(z) = \sqrt{f(z^2)} = z + b_3 z^3 + b_5 z^5 + \cdots,$$

used in proving Theorem 1.5 [cf. (1.19)]. As in the proof of Theorem 1.10, this function maps the disk $|z| < r < 1$ onto a domain D_r of area

$$\Sigma_r = \pi(r^2 + 3|b_3|^2r^6 + 5|b_5|^2r^{10} + \cdots).$$

Writing

$$\mu(r) = \max_{|z| \leqslant r} |\chi(z)|,$$

we have

$$\pi[\mu(r)]^2 \geqslant \Sigma_r = \pi(r^2 + 3|b_3|^2r^6 + 5|b_5|^2r^{10} + \cdots).$$

But according to Theorem 1.9,

$$[\mu(r)]^2 = \max_{|z| \leqslant r} |\chi(z)|^2 = \max_{|z| \leqslant r} |f(z^2)|$$
$$= \max_{|z| \leqslant r^2} |f(z)| \leqslant \frac{r^2}{(1 - r^2)^2}.$$

Therefore

$$\pi(r^2 + 3|b_3|^2r^6 + 5|b_5|^2r^{10} + \cdots) \leqslant \frac{\pi r^2}{(1 - r^2)^2},$$

which implies

$$r + 3|b_3|^2r^5 + 5|b_5|^2r^9 + \cdots \leqslant \frac{r}{(1 - r^2)^2}, \tag{1.37}$$

or

$$r^2 + |b_3|^2r^6 + |b_5|^2r^{10} + \cdots \leqslant \frac{r^2}{1 - r^2} \qquad (r < 1), \tag{1.38}$$

after integrating (1.37) from 0 to $r < 1$.

Next we observe that the left-hand side of (1.38) is just the integral

$$\frac{1}{2\pi}\int_0^{2\pi} |\chi(re^{i\theta})|^2 \, d\theta$$

[cf. (I.16.30)]. Clearly

$$\frac{1}{2\pi}\int_0^{2\pi} |\chi(re^{i\theta})|^2 \, d\theta = \frac{1}{2\pi}\int_0^{2\pi} |f(r^2e^{2i\theta})| \, d\theta = \frac{1}{4\pi}\int_0^{4\pi} |f(r^2e^{i\theta'})| \, d\theta',$$

where we have written $\theta' = 2\theta$ in the last integral. But

$$\int_{2\pi}^{4\pi} |f(r^2e^{i\theta'})| \, d\theta' = \int_0^{2\pi} |f(r^2e^{i\theta'})| \, d\theta',$$

since $f(r^2e^{i\theta'})$ is a periodic function of θ', and hence

$$\frac{1}{2\pi}\int_0^{2\pi} |\chi(re^{i\theta})|^2 \, d\theta = \frac{1}{2\pi}\int_0^{2\pi} |f(r^2e^{i\theta'})| \, d\theta'.$$

Therefore (1.38) can be written in the form

$$\frac{1}{2\pi}\int_0^{2\pi} |f(r^2e^{i\theta'})| \, d\theta' \leqslant \frac{r^2}{1 - r^2} \qquad (r < 1),$$

which reduces to (1.36), after replacing θ' by θ and r^2 by r.

THEOREM 1.12. *If the function*

$$f(z) = z + a_2z^2 + \cdots + a_nz^n + \cdots$$

is univalent on the unit disk $|z| < 1$, *then*

$$|a_n| < en \qquad (n \geqslant 2). \tag{1.39}$$

Proof. Since

$$a_n = \frac{1}{2\pi i}\int_{|z|=r} \frac{f(z)}{z^{n+1}}\,dz \qquad (r < 1)$$

[cf. (I.16.48)], it follows from Theorem 1.11 that

$$|a_n| \leqslant \frac{1}{r^n}\frac{1}{2\pi}\int_0^{2\pi} |f(re^{i\theta})|\,d\theta \leqslant \frac{1}{r^n}\frac{r}{1-r} = \frac{1}{r^{n-1}}\frac{1}{1-r}.$$

Setting

$$r = 1 - \frac{1}{n},$$

we find that

$$|a_n| \leqslant \frac{1}{\left(1 - \frac{1}{n}\right)^{n-1}} n = \left(1 + \frac{1}{n-1}\right)^{n-1} n < en,$$

as asserted.

Remark. The estimate (1.39) is not the best possible. We have already seen in Theorem 1.5 that a sharp estimate of $|a_2|$ (i.e., an estimate which cannot be improved) is given by $|a_2| \leqslant 2$. Similarly, it can be shown that $|a_3| \leqslant 3$ and $|a_4| \leqslant 4$ are sharp estimates. For higher coefficients, the best estimate presently available is due to Bazilevich, who has proved that

$$|a_n| \leqslant \tfrac{1}{2}en + 1.51 \qquad (n \geqslant 5),$$
$$\overline{\lim_{n\to\infty}} \frac{|a_n|}{n} \leqslant \frac{e}{2} = 1.3591\ldots$$

It has been conjectured by Bieberbach that

$$|a_n| \leqslant n \tag{1.40}$$

for all n. If this conjecture is true, then among all univalent functions of the form

$$f(z) = z + a_2z^2 + \cdots + a_nz^n + \cdots,$$

no function has coefficients with higher absolute values than the function

$$\frac{z}{(1 - e^{i\alpha}z)^2} = z + 2e^{i\alpha}z^2 + \cdots + ne^{(n-1)\alpha}z^n + \cdots.$$

For further results in the theory of univalent functions, we refer the reader to the books by Goluzin, Hayman, Jenkins and Nehari (cited in the Bibliography on pp. 351–353). The references and historical remarks in Jenkins' book are particularly detailed.

PROBLEMS

1.1. Can the finite z-plane with the origin deleted be mapped conformally onto the domain $0 < |w| < 1$?

1.2. Let G be an n-connected domain (see p. I.67), each of whose boundary continua contains more than one point. Prove that G can be mapped conformally onto a domain bounded by n closed Jordan curves.

1.3. Map the first quadrant $x > 0, y > 0$ onto the unit disk $\Delta: |w| < 1$. Map the "slit disk" $|z| < 1, z \notin [0, 1)$ onto Δ.

1.4. Let E be a continuum whose complement E^c is connected, and let $f(z)$ be an entire function taking no values in E. Prove that $f(z) \equiv \text{const}$.

Hint. Map E^c onto a disk.

1.5. Let G be the same as in Theorem 1.2, with conformal radius R_{z_0} ($z_0 \in G$). Prove that the circle $|z - z_0| = R_{z_0}$ always contains boundary points of G.

1.6. Let G and G^* be two domains, each satisfying the conditions of Theorem 1.2, and suppose $G \subset G^*$. Prove that $R_{z_0} \leqslant R^*_{z_0}$, where R_{z_0} and $R^*_{z_0}$ are the conformal radii of G and G^* relative to a point $z_0 \in G$.

Comment. In other words, enlargement of G leads to enlargement of its conformal radius.

1.7. Let G be the same as in Theorem 1.2, and let e_{z_0} be the family of all functions analytic and bounded by 1 on G satisfying the conditions

$$f(z_0) = 0, \quad f'(z_0) > 0 \qquad (z_0 \in G).$$

Prove that e_{z_0} contains a function $\varphi(z)$ such that $\varphi'(z_0)$ equals the largest value

$$\lambda_{z_0} = \sup_{f \in e_{z_0}} f'(z_0).$$

Prove that the function $w = \varphi(z)$ maps G conformally onto the unit disk $|w| < 1$. What is the relation between λ_{z_0} and the conformal radius of G?

1.8. Let $f(z) = u + iv$ be analytic on a domain G with area, and suppose $f(G)$ has area S. Prove that

$$S = \iint_G |f'(z)|^2 \, dx \, dy.$$

Hint. Write

$$S = \iint_G \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix} dx \, dy$$

and use the Cauchy-Riemann equations.

1.9. Let G be a simply connected domain whose boundary contains more than one point, and let $F(z)$ be any function analytic on G satisfying the conditions

$$F(z_0) = 0, \qquad F'(z_0) = 1 \qquad (z_0 \in G).$$

Prove that the integral

$$\iint_G |F'(z)|^2 \, dx \, dy$$

takes its minimum value for the function mapping G conformally onto the disk $|w| < R_{z_0}$, where R_{z_0} is the conformal radius of G. What is this minimum value?

1.10. According to Schwarz's lemma (Theorem I.17.8), if $f(z)$ is analytic and bounded by 1 on the unit disk $K: |z| < 1$ and if $f(0) = 0$, then $|f(z)| \leqslant |z|$ for all $z \in K$ and $|f'(0)| \leqslant 1$, where the equalities hold (the first for $z \neq 0$) if and only if $f(z) = e^{i\alpha} z$. Prove the following generalization of Schwarz's lemma, where the condition $f(0) = 0$ is dropped: If $f(z)$ is analytic and bounded by 1 on K, then

$$\left| \frac{f(\zeta) - f(z)}{1 - \overline{f(z)} f(\zeta)} \right| \leqslant \left| \frac{\zeta - z}{1 - \bar{z}\zeta} \right| \qquad (\zeta, z \in K),$$

$$|f'(z)| \leqslant \frac{1 - |f(z)|^2}{1 - |z|^2} \qquad (z \in K), \tag{1.41}$$

where the equalities hold if and only if

$$f(z) = e^{i\alpha} \frac{z + a}{1 + \bar{a}z} \qquad (a \in K).$$

Hint. The function

$$g(\zeta) = \frac{f\left(\dfrac{\zeta + z}{1 + \bar{z}\zeta}\right) - f(z)}{1 - \overline{f(z)} f\left(\dfrac{\zeta + z}{1 + \bar{z}\zeta}\right)}$$

satisfies the conditions of Schwarz's lemma.

1.11. Prove that the expression

$$\mu_G(z) = \frac{|f'(z)|}{1 - |f(z)|^2}$$

has the same value for every conformal mapping $w = f(z)$ of the domain G onto the unit disk $|w| < 1$.

Hint. Prove that $\mu_G(z)$ is invariant under the transformation

$$f_1(z) = e^{i\alpha} \frac{f(z) - a}{1 - \bar{a} f(z)} \qquad (|a| < 1).$$

1.12. Let G and D be two simply connected domains whose boundaries each contain more than one point. Let $f(z)$ be analytic on G, and suppose $f(G) \subset D$. Prove that

$$|f'(z)|\mu_D[f(z)] \leqslant \mu_G(z) \qquad (z \in G),$$

with the notation of Problem 1.11.

Hint. Use Schwarz's lemma.

1.13. Use the result of Problem 1.12 to deduce the inequality (1.41).

1.14. Let $f(z)$ be analytic on the disk $K: |z| < 1$, and suppose $f(z)$ takes no negative values on K. Prove that

$$\left|\frac{f'(z)}{f(z)}\right| \leqslant \frac{4}{1 - |z|^2},$$

$$\left(\frac{1 - |z|}{1 + |z|}\right)^2 \leqslant \left|\frac{f(z)}{f(0)}\right| \leqslant \left(\frac{1 + |z|}{1 - |z|}\right)^2$$

for all $z \in K$.

Hint. Use the result of Problem 1.12.

1.15. A domain G is said to be *convex* if every pair of points of G can be joined by a line segment contained in G (cf. Problems I.4.4 and I.13.2–I.13.5), and *starlike* with respect to a point $\zeta \in G$ if every point of G can be joined to ζ by a line segment contained in G. A closed Jordan curve L is said to be *convex* if $I(L)$ is convex, and *starlike* if $I(L)$ is starlike. Show that L is convex if and only if every straight line intersecting both $I(L)$ and $E(L)$ intersects L in just two points (cf. footnote 10, p. 20) and starlike with respect to $\zeta \in I(L)$ if and only if every ray emanating from ζ intersects L in only one point (cf. Theorem I.13.3). Prove that every convex domain is starlike (cf. Problem I.13.5).

1.16. The curve Γ figuring in the proof of Theorem 1.8 is convex if and only if the expression (1.30) is nonnegative. Write an analogous condition for Γ to be starlike with respect to the point $w = 0$.

Ans.

$$\frac{\partial \operatorname{Arg} f(z)}{\partial \theta} = \operatorname{Re}\left\{z\frac{f'(z)}{f(z)}\right\} \geqslant 0.$$

1.17. Let $f(z)$ be analytic in a neighborhood of the origin, and let r_1 be the largest value of r such that $f(z)$ is univalent on the disk $K_r: |z| < r$. Similarly, let r_2 be the largest value of r such that $f(z)$ is *convex univalent* on K_r, i.e., $f(z)$ is univalent on K_r and $f(K_r)$ is convex, and let r_3 be the largest value of r such that $f(z)$ is *starlike univalent* on K_2, i.e., $f(z)$ is univalent on K_r and $f(K_r)$ is starlike with respect to $w = 0$. Then obviously $r_2 \leqslant r_3 \leqslant r_1$. Find the values of r_1, r_2, and r_3 for each of the following functions:

a) $\dfrac{z}{1 - z}$; b) $z + az^2$ (a real); c) $\dfrac{z}{1 - z^2}$.

Ans.

b) $r_1 = \dfrac{1}{2|a|}$, $r_2 = \dfrac{1}{4|a|}$, $r_3 = \dfrac{1}{2|a|}$.

1.18. Let $f(z)$ be analytic on a convex domain G, and suppose $\operatorname{Re} f'(z) > 0$ for all $z \in G$. Prove that $f(z)$ is univalent on G.

Hint. Calculate

$$f(z_2) - f(z_1) = \int_L f'(z)\, dz,$$

where L is the line segment joining z_1 to z_2.

1.19. Let G be the same as in Theorem 1.2, and let $w = F(z)$ be the function mapping G conformally onto the disk $|w| < R_{z_0}$ and satisfying the conditions

$$F(z_0) = 0, \quad F'(z_0) = 1 \qquad (z_0 \in G),$$

where R_{z_0} is the conformal radius of G. Let ρ_{z_0} denote the distance between z_0 and the boundary of G. Prove that

$$\rho_{z_0} \leqslant R_{z_0} \leqslant 4\rho_{z_0}.$$

Hint. Use Theorems 1.6 and 1.10.

Comment. Thus $R_{z_0} \to 0$ as $\rho_{z_0} \to 0$.

1.20. Let

$$f(z) = \sum_{n=1}^{\infty} a_n z^n$$

be convex univalent on the unit disk $K: |z| < 1$, let

$$g(z) = \sum_{n=1}^{\infty} b_n z^n$$

be analytic on K, and suppose $g(K) \subset f(K)$. Prove that $|b_n| \leqslant |a_1|$ and in particular $|a_n| \leqslant |a_1|$ for $n \geqslant 1$.

Hint. Use Schwarz's lemma and the fact that

$$G(z) = \frac{1}{m} \sum_{n=1}^{m} g(\varepsilon_n z^{1/m}) = b_m z^m + b_{2m} z^{2m} + \cdots,$$

where ε_n $(n = 1, \ldots, m)$ are the mth roots of unity.

1.21. Suppose

$$f(z) = z + a_2 z^2 + \cdots + a_n z^n + \cdots$$

is convex univalent on the unit disk $|z| < 1$. Prove that

$$|a_n| \leqslant 1 \qquad (n \geqslant 2),$$

where the estimate is sharp.

1.22. Suppose

$$f(z) = z + a_2 z^2 + \cdots + a_n z^n + \cdots$$

is starlike univalent on the unit disk $|z| < 1$. Prove that

$$|a_n| \leqslant n \qquad (n \geqslant 2),$$

where the estimate is sharp.

Hint. Use the results of Problems 1.20 and 1.21.

1.23. Given a function $f(z)$ analytic on the unit disk $K: |z| < 1$, suppose $f(z)$ is real if and only if $z \in K$ is real. Then $f(z)$ is said to be *typically real* (on K). Prove that if

$$f(z) = z + a_2 z^2 + \cdots + a_n z^n + \cdots$$

is typically real, then

$$|a_n| \leqslant n \qquad (n \geqslant 2),$$

where the estimate is sharp.

Hint. $\operatorname{Im} f(z)$ has constant sign for $z \in K$, $\operatorname{Im} z > 0$.

Comment. Note that if $f(z)$ is univalent on K, with real coefficients, then $f(z)$ is typically real.

2

CONFORMAL MAPPING: RAMIFICATIONS

4. Conformal Mapping of Sequences of Domains

Let $\{G_n\}$ be a sequence of simply connected domains in the z-plane, each containing a fixed disk k with center z_0. Let E be the set of all points z with a neighborhood $\mathcal{N}(z)$ contained in all the domains G_n starting from some value of n (depending on z). Obviously E is nonempty, since $k \subset E$. Moreover E is open, and hence, according to Theorem I.4.7, E is the union of countably many disjoint domains, namely the (connected) components of E. Of these components, let G_{z_0} be the one containing z_0 (and hence k). Then G_{z_0} is called the *kernel* of the sequence $\{G_n\}$ (relative to the point z_0).

Remark 1. If F is a compact subset of the kernel G_{z_0}, then F is contained in all the domains G_n starting from some value of n. In fact, if $F \subset G_{z_0}$, then for every point $z \in F$, we can find a neighborhood $\mathcal{N}(z)$ such that $\mathcal{N}(z) \subset G_n$ for all $n > \nu(z)$. Therefore, according to the Heine-Borel theorem (Theorem I.3.14), there exists a finite number of points $z_1, \ldots, z_m$ in F such that

$$F \subset \mathcal{N}(z_1) \cup \cdots \cup \mathcal{N}(z_m).$$

Then, if

$$\nu = \max\{\nu(z_1), \ldots, \nu(z_m)\},$$

all the neighborhoods $\mathcal{N}(z_1), \ldots, \mathcal{N}(z_m)$ are contained in G_n if $n > \nu$, and hence the same is true of the set F.

Remark 2. The kernel G_{z_0} of the sequence $\{G_n\}$ can be characterized as the largest domain containing z_0 such that every compact subset $F \subset G_{z_0}$ belongs to all the G_n starting from some value of n (depending on F). In

fact, if g is a domain with this property, then every point $z \in g$ belongs to all the G_n starting from some value of n (depending on z). Therefore $g \subset E$, which means that if $z_0 \in g$, then g is contained in the component of E containing z_0, i.e., $g \subset G_{z_0}$, as asserted.

DEFINITION. *Let $\{G_n\}$ be a sequence of simply connected domains, with kernel G_{z_0} relative to the point z_0. Then $\{G_n\}$ is said to converge to G_{z_0} if every subsequence of $\{G_n\}$ has the same kernel G_{z_0} (relative to z_0) as $\{G_n\}$ itself. Otherwise, $\{G_n\}$ is said to diverge.*

Example 1. Let G_n $(n = 1, 2, \ldots)$ be the domain shown in Figure 2.1, consisting of two fixed rectangles D' and D'' joined by another rectangle δ_n of constant length l and height h_n, where h_n decreases monotonically to zero. Then the set E figuring in the construction of G_{z_0} consists of the domains D' and D''. The kernel of the sequence $\{G_n\}$ relative to any point $z_0' \in D'$ is the rectangle D', while the kernel of $\{G_n\}$ relative to any point $z_0'' \in D''$ is the rectangle D''. Moreover, $\{G_n\}$ converges to the kernel D' (relative to z_0') and also to the kernel D'' (relative to z_0'').

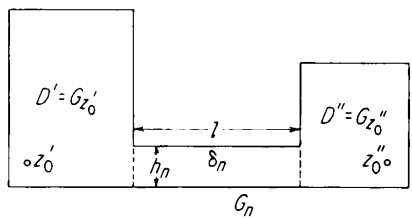


FIGURE 2.1

Example 2. Let $\{G_n\}$ be the sequence of domains illustrated schematically in Figure 2.2, where every domain G_{2m-1} with odd index coincides with the same rectangle D', while every domain G_{2m} with even index coincides with another rectangle D'', which, although different from D', intersects D' in a third rectangle δ. Then the kernel of the sequence $\{G_n\}$ relative to any point $z_0 \in \delta$ is the rectangle δ. However, the kernel of the subsequence $\{G_{2m-1}\}$ relative to any point $z_0 \in \delta$ is D', while the kernel of the subsequence $\{G_{2m}\}$ relative to z_0 is D''. Therefore, in this case, the sequence $\{G_n\}$ diverges.

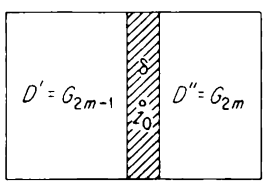


FIGURE 2.2

The relevance of these concepts to the theory of conformal mapping is revealed by the following theorem due to Carathéodory, on the mapping of domains "with variable boundaries":

THEOREM 2.1 (*Carathéodory's mapping theorem*). *Let $\{G_n\}$ be a sequence of simply connected domains contained in a fixed disk $K: |z| < R$,*[1] *and containing another fixed disk $k: |z - z_0| < \rho$. For every $n = 1, 2, \ldots$, let $w = f_n(z)$ be the function which maps G_n conformally onto the unit disk $\Delta: |w| < 1$ and satisfies the conditions*

$$f_n(z_0) = 0, \qquad f_n'(z_0) > 0,$$

[1] In other words, the domains $G_1, \ldots, G_n, \ldots$ are uniformly bounded.

and let $z = \varphi_n(w)$ *be the inverse of* $w = f_n(z)$. *Moreover, let* $w = f(z)$ *be the function which maps the kernel* G_{z_0} *of the sequence* $\{G_n\}$ *(relative to the point* z_0*) conformally onto* Δ *and satisfies the conditions*

$$f(z_0) = 0, \qquad f'(z_0) > 0,$$

and let $z = \varphi(w)$ *be the inverse of* $w = f(z)$.[2] *Under these conditions, if the sequence of domains* $\{G_n\}$ *converges to* G_{z_0}, *then the sequence of functions* $\{f_n(z)\}$ *converges uniformly inside* G_{z_0} *to the function* $f(z)$, *and the sequence* $\{\varphi_n(w)\}$ *converges uniformly inside* Δ *to the function* $\varphi(w)$. *Conversely, if* $\{f_n(z)\}$ *converges uniformly inside* G_{z_0} *to* $f(z)$, *or if* $\{\varphi_n(w)\}$ *converges uniformly inside* Δ *to* $\varphi(w)$, *then* $\{G_n\}$ *converges to* G_{z_0}.

Proof. For the time being, we make no assumptions about the convergence of the sequence of domains $\{G_n\}$. Every function $f_n(z)$ is defined on its own domain G_n, and it may happen that there are points in G_{z_0} at which any given function $f_n(z)$ is undefined. For example, this is the case for an increasing sequence of domains $\{G_n\}$, $\overline{G}_n \subset G_{n+1}$ $(n = 1, 2, \ldots)$ approximating a domain G which is the kernel of $\{G_n\}$, as shown in Figure 2.3.[3] However, every closed set $F \subset G_{z_0}$ (in particular, every closed disk contained in G_{z_0}) is contained in all the domains G_n starting from some value of n, and therefore every function $f_n(z)$ is defined and analytic on F starting from some value of n. Since all the concepts and theorems pertaining to compact families of analytic functions deal only with the behavior of functions on closed subsets of a given domain, the theory of compact families is applicable to the present case, i.e., to the sequence $\{f_n(z)\}$ and the domain G_{z_0}.

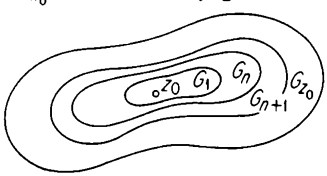


FIGURE 2.3

Now let $\{n_k\}$ be any sequence of positive integers such that $n_1 < \cdots < n_k < \cdots$, and let G'_{z_0} be the kernel of the sequence of domains $\{G_{n_k}\}$. The functions $f_{n_k}(z)$, $k = 1, 2, \ldots$ are uniformly bounded inside G'_{z_0} (in fact $|f_{n_k}(z)| < 1$), and hence, by Montel's compactness criterion (Theorem I.17.17), $\{n_k\}$ contains a subsequence $\{n'_k\}$ such that the sequence $\{f_{n'_k}(z)\}$ is uniformly convergent inside G'_{z_0}. Moreover, the functions $\{\varphi_{n'_k}(w)\}$, $k = 1, 2, \ldots$ are uniformly bounded on the unit disk $\Delta: |w| < 1$ (in fact $|\varphi_{n'_k}(w)| < R$), and hence from $\{n'_k\}$ we can select a new subsequence $\{\tilde{n}_k\}$ such that the sequence $\{\varphi_{\tilde{n}_k}(w)\}$ is uniformly convergent inside Δ. Obviously, the sequence $\{f_{\tilde{n}_k}(z)\}$ is uniformly convergent inside G'_{z_0}, as before. The kernel $\tilde{G}_{z_0}$ of the sequence of domains

[2] Here it is tacitly assumed that G_{z_0} is simply connected (however, see Problem 2.1).

[3] The terminology resembles that on p. I.373, and on p. 81 below.

$\{G_{\tilde{n}_k}\}$ may differ from G'_{z_0}, but in any event $\tilde{G}_{z_0} \supset G'_{z_0}$, since $\{G_{\tilde{n}_k}\}$ is a subsequence of $\{G_{n'_k}\}$.

Since the sequence $\{f_{\tilde{n}_k}(z)\}$ is uniformly bounded and convergent inside $G'_{z_0} \subset \tilde{G}_{z_0}$, it follows from Vitali's theorem (Theorem I.17.18) that $\{f_{\tilde{n}_k}(z)\}$ is uniformly convergent inside $\tilde{G}_{z_0}$. Thus we have proved that the original increasing sequence of positive integers $\{n_k\}$ contains a subsequence $\{\tilde{n}_k\}$ such that the sequence of functions $\{f_{\tilde{n}_k}(z)\}$ converges uniformly inside the kernel $\tilde{G}_{z_0}$ of the sequence of domains $\{G_{\tilde{n}_k}\}$ to an analytic function $\tilde{f}(z)$, while at the same time, the sequence of inverse functions $\{\varphi_{\tilde{n}_k}(w)\}$ converges uniformly inside the unit disk Δ to an analytic function $\tilde{\varphi}(w)$. Obviously

$$\tilde{f}(z_0) = \lim_{\tilde{n}_k \to \infty} f_{\tilde{n}_k}(z_0) = 0, \qquad \tilde{f}'(z_0) = \lim_{\tilde{n}_k \to \infty} f'_{\tilde{n}_k}(z_0) \geqslant 0,$$

and similarly

$$\tilde{\varphi}(0) = \lim_{\tilde{n}_k \to \infty} \varphi_{\tilde{n}_k}(0) = z_0, \qquad \tilde{\varphi}'(0) = \lim_{\tilde{n}_k \to \infty} \varphi'_{\tilde{n}_k}(0) \geqslant 0.$$

Thus both the numbers $f'_{\tilde{n}_k}(z_0)$, $k = 1, 2, \ldots$ and their reciprocals

$$\varphi'_{\tilde{n}_k}(0) = \frac{1}{f'_{\tilde{n}_k}(z_0)} \qquad (k = 1, 2, \ldots)$$

have finite limits. But then both limits must be nonzero, i.e., $\tilde{f}'(z_0) > 0$ and $\tilde{\varphi}'(0) > 0$, where

$$\tilde{f}'(z_0)\tilde{\varphi}'(0) = 1. \tag{2.1}$$

In particular, $\tilde{f}(z) \not\equiv \text{const}$, $\tilde{\varphi}(w) \not\equiv \text{const}$, and hence, according to Theorem II.4.4, $\tilde{f}(z)$ and $\tilde{\varphi}(w)$ are univalent on the domains $\tilde{G}_{z_0}$ and Δ, respectively. Clearly, $\tilde{f}(\tilde{G}_{z_0})$ contains the point 0. Moreover, since $|f_{\tilde{n}_k}(z)| < 1$, it follows that $|\tilde{f}(z)| \leqslant 1$ and hence that $\tilde{f}(\tilde{G}_{z_0}) \subset \Delta$. Similarly, $\tilde{\varphi}(\Delta)$ contains the point z_0, and as we now show, $\tilde{\varphi}(\Delta)$ is contained in $\tilde{G}_{z_0}$.

With this in mind, let $\tilde{w} \in \Delta$, so that $\tilde{z} = \tilde{\varphi}(w)$ is a point of the domain $\tilde{\varphi}(\Delta)$. Then it is not hard to see that $\tilde{z}$ has a neighborhood contained in every domain $G_{\tilde{n}_k}$ starting from some value of $\tilde{n}_k$. In fact, the function $z = \tilde{\varphi}(w)$ maps any circle $\delta: |w - \tilde{w}| = \tilde{\rho}$ contained in Δ into a closed Jordan curve γ contained in $\tilde{\varphi}(\Delta)$, and moreover, $I(\gamma) \subset \tilde{\varphi}(\Delta)$, $\tilde{z} \in I(\gamma)$, where $I(\gamma)$ denotes the interior of γ (cf. Theorem II.4.5). Let $C: |z - \tilde{z}| \leqslant \tilde{r}$ be a closed disk contained in $I(\gamma)$, and let ε be the distance between C and γ. Then obviously $\varepsilon > 0$, and

$$|\tilde{\varphi}(w) - \zeta| \geqslant \varepsilon$$

if $w \in \delta$ and $\zeta \in C$. On the other hand, the inequality

$$|\varphi_{\tilde{n}_k}(w) - \tilde{\varphi}(w)| < \varepsilon$$

holds for all $w \in \delta$ and sufficiently large $\tilde{n}_k > N$. Therefore, according to Rouché's theorem (Theorem II.2.4), the functions $\tilde{\varphi}(w) - \zeta$ and

$$\varphi_{\tilde{n}_k}(w) - \zeta = [\tilde{\varphi}(w) - \zeta] + [\varphi_{\tilde{n}_k}(w) - \tilde{\varphi}(w)]$$

have the same number of zeros inside γ, i.e., every point $\zeta \in C$ belongs to the domain $\varphi_{\tilde{n}_k}(\Delta)$ if $\tilde{n}_k > N$. In other words, every point $\tilde{z} \in \tilde{\varphi}(\Delta)$ has a neighborhood (the open disk $|z - \tilde{z}| < \tilde{r}$) contained in all the domains $G_{\tilde{n}_k}$ starting from some value of $\tilde{n}_k$. Since $\tilde{\varphi}(\Delta)$ is a domain containing z_0, it follows from the definition of the kernel of the sequence $\{G_{\tilde{n}_k}\}$ that $\tilde{\varphi}(\Delta) \subset \tilde{G}_{z_0}$, as asserted.

Thus $z = \tilde{\varphi}(w)$ maps Δ conformally onto the domain $\tilde{\varphi}(\Delta) \subset \tilde{G}_{z_0}$, while $w = \tilde{f}(z)$ maps $\tilde{G}_{z_0}$ conformally onto the domain $\tilde{f}(\tilde{G}_{z_0}) \subset \Delta$. Therefore the function $\psi(w) = \tilde{f}[\tilde{\varphi}(w)]$ maps Δ conformally onto the domain $\tilde{f}(\tilde{G}_{z_0}) \subset \Delta$. Noting that

$$\psi(0) = \tilde{f}[\tilde{\varphi}(0)] = \tilde{f}(z_0) = 0, \qquad \psi'(0) = \tilde{f}'(z_0)\tilde{\varphi}'(0) = 1$$

[cf. (2.1)], we conclude from Schwarz's lemma (Theorem I.17.8) that $\psi(w) = e^{i\alpha}w$, where $e^{i\alpha} = \psi'(0) = 1$. It follows that

$$\psi(w) = \tilde{f}[\tilde{\varphi}(w)] = w \quad \text{or} \quad \tilde{\varphi}(w) = \tilde{f}^{-1}(w),$$

i.e., $\tilde{\varphi}(w)$ is the inverse of the function $\tilde{f}(z)$. Therefore, if w is an arbitrary point of the disk Δ, then $z = \tilde{\varphi}(w) \in \tilde{\varphi}(\Delta) \subset \tilde{G}_{z_0}$ and $\tilde{f}(z) = w \in \tilde{f}(\tilde{G}_{z_0})$, i.e., $\tilde{f}(\tilde{G}_{z_0}) \supset \Delta$. But as shown above, $\tilde{f}(\tilde{G}_{z_0}) \subset \Delta$, and hence $\tilde{f}(\tilde{G}_{z_0}) = \Delta$. In just the same way, we see that $\tilde{\varphi}(\Delta) = \tilde{G}_{z_0}$. Thus the functions $w = \tilde{f}(z)$ and $z = \tilde{\varphi}(w)$ are inverses of each other, and $w = \tilde{f}(z)$ maps the kernel $\tilde{G}_{z_0}$ of the sequence $\{G_{n_k}\}$ conformally onto the unit disk, while satisfying the conditions

$$\tilde{f}(z_0) = 0, \qquad \tilde{f}'(z_0) > 0. \tag{2.2}$$

Having established this fact, we are now in a position to complete the proof. We begin by proving the first assertion of the theorem. Suppose the sequence of domains $\{G_n\}$ converges to its kernel G_{z_0}. Then the kernel remains the same for every subsequence $\{G_{\tilde{n}_k}\}$, i.e., $\tilde{G}_{z_0} = G_{z_0}$. Moreover, the limit function

$$\tilde{f}(z) = \lim_{\tilde{n}_k \to \infty} f_{\tilde{n}_k}(z)$$

maps G_{z_0} conformally onto the unit disk $\Delta: |w| < 1$ and satisfies the conditions (2.2). It follows from the uniqueness theorem for conformal mapping that all possible uniformly convergent subsequences $\{f_{\tilde{n}_k}(z)\}$ have the same limit $\tilde{f}(z) \equiv f(z)$. We now show that the original sequence $\{f_n(z)\}$ converges uniformly to $f(z)$ inside G_{z_0}. Suppose this is not the case. Then we can find a closed set $F \subset G_{z_0}$, a sequence $\{n_k\}$ of positive

integers (where $n_1 < \cdots < n_k < \cdots$) and a sequence $\{z_k\}$ of points in F, such that

$$|f_{n_k}(z_k) - f(z_k)| \geqslant \alpha > 0 \qquad (k = 1, 2, \ldots). \tag{2.3}$$

But because of the compactness, the sequence $\{f_{n_k}(z)\}$ must contain a subsequence $f_{n_k^*}(z)$ which converges uniformly inside G_{z_0}, in particular on the set F, to the function $f(z)$. It follows that

$$|f_{n_k^*}(z) - f(z)| < \alpha$$

for all $z \in F$ and sufficiently large $n_k^* > N$. In particular, we have

$$|f_{n_k^*}(z_k) - f(z_k)| < \alpha \tag{2.4}$$

for all $n_k^* > N$. But (2.4) contradicts (2.3), and hence our assumption that $\{f_n(z)\}$ does not converge uniformly to $f(z)$ inside G_{z_0} must be false.[4]

Finally, to prove the second assertion of the theorem, suppose $\{f_n(z)\}$ converges uniformly to $f(z)$ inside G_{z_0}, but $\{G_n\}$ does not converge to its kernel G_{z_0}. Then there exists a subsequence $\{G_{n_k}\}$ whose kernel is different from G_{z_0} (and contains G_{z_0} as a proper subset). From $\{f_{n_k}(z)\}$ we select a new subsequence $\{f_{\tilde{n}_k}(z)\}$ such that $\{f_{\tilde{n}_k}(z)\}$ is uniformly convergent inside the kernel $\tilde{G}_{z_0}$ of the corresponding sequence of domains $\{G_{\tilde{n}_k}\}$ and such that the sequence of inverse functions $\{\varphi_{\tilde{n}_k}(w)\}$ is uniformly convergent inside the unit disk $\Delta: |w| < 1$. Then, just as before, the limit function

$$\tilde{f}(z) = \lim_{\tilde{n}_k \to \infty} f_{\tilde{n}_k}(z)$$

maps the domain $\tilde{G}_{z_0}$ conformally onto Δ. Therefore $\tilde{f}(z)$ cannot coincide with the function $f(z)$ which maps G_{z_0} conformally onto Δ, since G_{z_0} is a proper subset of $\tilde{G}_{z_0}$. But this is impossible, since $\{f_{\tilde{n}_k}(z)\}$ is a subsequence of the uniformly convergent sequence $\{f_n(z)\}$. This contradiction shows that $\{G_n\}$ must converge to G_{z_0}, thereby completing the proof of the theorem (except for some details left as an exercise for the reader).

Example. Let G_n, z_0', z_0'', D' and D'' have the same meaning as in Example 1, p. 33. Moreover, let $f_n(z)$ and $F_n(z)$ be the functions mapping G_n conformally onto the unit disk and satisfying the conditions

$$f_n(z_0') = 0, \qquad f_n'(z_0') > 0$$

and

$$F_n(z_0'') = 0, \qquad F_n'(z_0'') > 0,$$

[4] The same argument has already been encountered in the proof of Vitali's theorem (Theorem I.17.18).

respectively. Then the sequence $\{f_n(z)\}$ converges uniformly inside D' to a function $f(z)$ mapping D' conformally onto the unit disk, while $\{F_n(z)\}$ converges uniformly inside D'' to a function $F(z)$ mapping D'' conformally onto the unit disk.

We now consider sequences of domains of a particularly simple kind: A sequence of domains $\{G_n\}$ is said to be *decreasing* if $\overline{G}_{n+1} \subset G_n$ for all $n = 1, 2, \ldots$ (Obviously, if G_1 is bounded, the domains $G_1, \ldots, G_n, \ldots$ are uniformly bounded.) By a *Jordan domain*, we mean a domain which is the interior of a closed Jordan curve.

THEOREM 2.2. *If G is a Jordan domain, then there exists a decreasing sequence of bounded simply connected domains $\{G_n\}$ converging to G as its unique kernel.*

Proof. Given any integer $n > 0$, let G_n' be the set of all points whose distance from $\overline{G}$ is less than $1/n$. Obviously, G_n' is open and connected (and hence a domain), and moreover $\overline{G} \subset G_n'$. Among the components of $(\overline{G}_n')^c$, the complement of $\overline{G}_n'$, one domain $G_{n,\infty}'$ contains the point at infinity, while the others are bounded, with boundaries contained in the boundary of G_n'. Let G_n denote the set consisting of G_n', all the bounded components of $(\overline{G}_n')^c$ and all the boundary points of G_n' which are not boundary points of $G_{n,\infty}'$. Then $G_n \supset G_n'$ is simply connected, since its boundary coincides with that of $G_{n,\infty}'$ and hence is a continuum (see Problem 2.3). The sequence of domains $\{G_n\}$ clearly satisfies the conditions

$$\overline{G} \subset G_{n+1} \subset \overline{G}_{n+1} \subset G_n \qquad (n = 1, 2, \ldots).$$

Therefore $\{G_n\}$ is decreasing, and G must be contained in some kernel of $\{G_n\}$. But no point z' in the exterior of G can belong to this (or any other) kernel of $\{G_n\}$. In fact, let γ be an arc which joins z' to the point at infinity and lies in the exterior of G, and let $\rho > 0$ be the distance between $\overline{G}$ and γ. Then $\gamma \subset G_{n,\infty}'$, and hence $z' \notin G_n$ if $n > 1/\rho$, i.e., z' does not belong to a kernel of the sequence $\{G_n\}$. It follows that G is the unique kernel of $\{G_n\}$. Moreover, $\{G_n\}$ converges to G, since the same argument applies to any subsequence $\{G_{n_k}\}$.

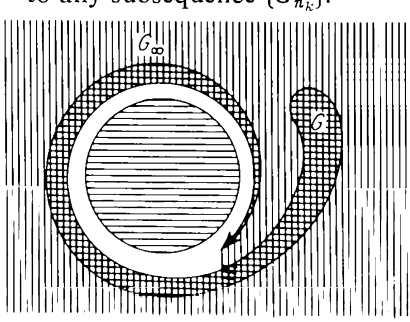


FIGURE 2.4

DEFINITION. *Let G be a bounded simply connected domain, and let G_∞ be the component of $(\overline{G})^c$ containing the point at infinity. Then G is said to be a Carathéodory domain if G and G_∞ have the same boundary.*

In particular, every Jordan domain is a Carathéodory domain. Figure 2.4 shows a Carathéodory domain G whose

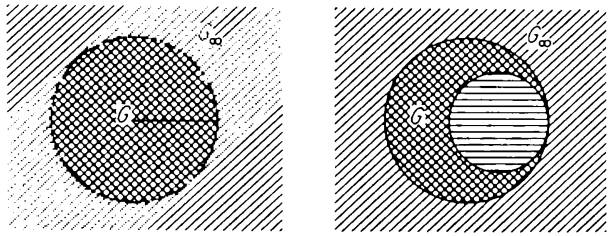


FIGURE 2.5

boundary separates the plane into three distinct domains (therefore G is not a Jordan domain), while Figure 2.5 shows two domains which are not Carathéodory domains.

THEOREM 2.3. *If G is a Carathéodory domain, then there exists a decreasing sequence of bounded simply connected domains $\{G_n\}$ converging to G as its kernel (relative to any point $z_0 \in G$).*

Proof. The construction of the sequence $\{G_n\}$ is the same as in the proof of Theorem 2.2, except that now we can no longer assert that no exterior point of G belongs to a kernel of $\{G_n\}$, but only that no point of G_∞ belongs to a kernel of $\{G_n\}$. However, of the (in general) several kernels of $\{G_n\}$, there is one which coincides with G, i.e., the kernel G_{z_0} relative to any point $z_0 \in G$. In fact, since $z_0 \in G$ and $G \subset G_n$ ($n = 1, 2, \ldots$) by construction, we have $G \subset G_{z_0}$. On the other hand, no point z' in the boundary of G can belong to G_{z_0}, since, by definition, z' also belongs to the boundary of G_∞, so that every neighborhood of z' contains points of G_∞. But G_{z_0} can contain no points of G_∞, and hence no boundary points of G, i.e., $G \supset G_{z_0}$. It follows that $G = G_{z_0}$, as asserted.

The following result will be needed later in our study of polynomial approximation (see Sec. 16):

THEOREM 2.4. *Let G be a kernel of a decreasing sequence of bounded simply connected domains $\{G_n\}$. For every $n = 1, 2, \ldots$, let $w = f_n(z)$ be the function which maps G_n conformally onto the unit disk $|w| < 1$ and satisfies the conditions*

$$f_n(z_0) = 0, \quad f_n'(z_0) > 0 \qquad (z_0 \in G).$$

Moreover, let $w = f(z)$ be the function which maps G conformally onto the unit disk and satisfies the conditions

$$f(z_0) = 0, \qquad f'(z_0) > 0.$$

Then

$$\lim_{n\to\infty} \iint_G |[f_n(z)]^k f_n'(z) - [f(z)]^k f'(z)|^2 \, dx\, dy = 0 \tag{2.5}$$

for every integer $k \geqslant 0$.[5]

Proof. Using the elementary inequality $|a+b|^2 \leqslant 2|a|^2 + 2|b|^2$ and the fact that $|f_n(z)| < 1$, $|f(z)| < 1$, we write

$$\begin{aligned} &|[f_n(z)]^k f_n'(z) - [f(z)]^k f'(z)|^2 \\ &\qquad \leqslant 2|f_n(z)|^{2k}|f_n'(z) - f'(z)|^2 + 2|f'(z)|^2|[f_n(z)]^k - [f(z)]^k|^2 \\ &\qquad < 2|f_n'(z) - f'(z)|^2 + 2k^2|f'(z)|^2|f_n(z) - f(z)|^2. \end{aligned}$$

It follows that

$$\begin{aligned} &\iint_G |[f_n(z)]^k f_n'(z) - [f(z)]^k f'(z)|^2 \, dx\, dy \\ &\qquad < 2\iint_G |f_n'(z) - f'(z)|^2 \, dx\, dy + 2k^2 \iint_G |f'(z)|^2 |f_n(z) - f(z)|^2 \, dx\, dy. \end{aligned} \tag{2.6}$$

We now proceed to estimate the two integrals in the right-hand side of (2.6).

Let C_r denote the inverse image of the circle $|w| = r < 1$ under the mapping $w = f(z)$, and let g_r denote the interior of C_r. According to Theorem 2.1, the sequences $\{f_n(z)\}$ and $\{f_n'(z)\}$ converge uniformly inside G to the functions $f(z)$ and $f'(z)$, respectively (see also Problem 2.6 and Theorem I.15.8). Given any $\varepsilon > 0$, we first choose numbers r_0 and $\rho_0 < r_0$ in the interval $(0, 1)$ such that

$$\pi - \pi\rho_0^2 < \varepsilon^2, \qquad r_0 - \rho_0 < \varepsilon,$$

and then an integer $N(\varepsilon)$ such that

$$|f_n(z) - f(z)| < r_0 - \rho_0, \qquad |f_n'(z) - f'(z)| < \varepsilon$$

for all $z \in C_{r_0}$ (and hence for all $z \in g_{r_0}$) if $n > N(\varepsilon)$. It follows that

$$\begin{aligned} &\iint_G |f_n'(z) - f'(z)|^2 \, dx\, dy \\ &\quad = \iint_{g_{r_0}} |f_n'(z) - f'(z)|^2 \, dx\, dy + \iint_{G - g_{r_0}} |f_n'(z) - f'(z)|^2 \, dx\, dy \\ &\quad < \varepsilon^2 \cdot \text{Area of } g_{r_0} + 2\iint_{G - g_{r_0}} |f_n'(z)|^2 \, dx\, dy + 2\iint_{G - g_{r_0}} |f'(z)|^2 \, dx\, dy. \end{aligned} \tag{2.7}$$

[5] Concerning the rigorous interpretation of double integrals of the type (2.5), see the proof of Theorem 1.10 and footnote 25, p. 116.

The integral

$$\iint_{G-g_{r_0}} |f'(z)|^2 \, dx \, dy$$

represents the area of the image of $G \cap E(C_{r_0})$ under the mapping $w = f(z)$ [cf. Problem 1.8].[6] This is just the area of the annulus $r_0 < |w| < 1$, which equals $\pi - \pi r_0^2$ and is less than ε^2 (by the choice of r_0). The integral

$$\iint_{G-g_{r_0}} |f_n'(z)|^2 \, dx \, dy$$

represents the area of the image of the same set $G \cap E(C_{r_0})$, but under the mapping $w = f_n(z)$.[7] Since $|f_n(z) - f(z)| < r_0 - \rho_0$ for all $z \in C_{r_0}$, this image lies between the circle $|w| = \rho_0$ and the unit circle, and hence is less than $\pi - \pi\rho_0^2 < \varepsilon^2$. Therefore (2.7) implies

$$\iint_G |f_n'(z) - f'(z)|^2 \, dx \, dy < \varepsilon^2(S + 4), \tag{2.8}$$

since

$$\text{Area of } g_{r_0} < \text{Area of } G = S.$$

Similarly, turning to the second integral in the right-hand side of (2.6), we find that

$$\begin{aligned}
&\iint_G |f'(z)|^2 |f_n(z) - f(z)|^2 \, dx \, dy \\
&\quad = \iint_{g_{r_0}} |f'(z)|^2 |f_n(z) - f(z)|^2 \, dx \, dy + \iint_{G-g_{r_0}} |f'(z)|^2 |f_n(z) - f(z)|^2 \, dx \, dy \\
&\quad < \varepsilon^2 \iint_{g_{r_0}} |f'(z)|^2 \, dx \, dy + 4 \iint_{G-g_{r_0}} |f'(z)|^2 \, dx \, dy < \pi\varepsilon^2 + 4\varepsilon^2.
\end{aligned} \tag{2.9}$$

Therefore, combining (2.6), (2.8) and (2.9), we obtain

$$\iint_G |[f_n(z)]^k f_n'(z) - [f(z)]^k f'(z)|^2 \, dx \, dy < \varepsilon^2[2S + 8 + 2k^2(\pi + 4)]$$

if $n > N(\varepsilon)$, which is equivalent to (2.5), since the expression on the right can be made arbitrarily small.

---

[6] $E(C_{r_0})$ denotes the exterior of C_{r_0}.

[7] Here it is important to note that every G_n contains $G \cap E(C_{r_0})$.

5. Curvilinear Half-Intervals

Riemann's mapping theorem (Theorem 1.2) tells us that any simply connected domain with a boundary Γ consisting of more than one point can be mapped conformally onto the unit disk, but it contains no information about the behavior of the mapping on Γ itself. We now turn to the important problem of the boundary behavior of conformal mappings (see esp. Sec. 8), after first introducing the key notion of an accessible boundary point in Sec. 6.

By a *half-open curvilinear interval*, or briefly, a *curvilinear half-interval*, we mean the one-to-one continuous image γ of a half-open interval $[\alpha, \beta)$. Thus if γ has the equation

$$z = \lambda(t), \qquad \alpha \leqslant t < \beta,$$

then $\lambda(t') \neq \lambda(t'')$ if $t' \neq t''$. The point $z = \lambda(\alpha)$ is called the *initial point* of γ, and the set E_γ of limit points of all possible sequences $\{z_n\} = \{\lambda(t_n)\}$, $\alpha \leqslant t_n < \beta$, where $t_n \to \beta -$ as $n \to \infty$, is called the *limit set* of γ. Two curvilinear half-intervals $\gamma: z = \lambda(t)$, $\alpha \leqslant t < \beta$ and $\gamma: z = \tilde{\lambda}(\tilde{t})$, $\tilde{\alpha} \leqslant \tilde{t} < \tilde{\beta}$ are regarded as identical if there exists a continuous strictly monotonic function $\tilde{t} = \mu(t)$ transforming $[\alpha, \beta)$ into $[\tilde{\alpha}, \tilde{\beta})$ and $\lambda(t)$ into $\tilde{\lambda}(\tilde{t})$ [cf. p. I.60].

Example 1. Any Jordan arc minus its final point Z is a curvilinear half-interval,[8] with a limit set consisting of the single point Z. Obviously, any curvilinear half-interval with a limit set consisting of a single point Z can be transformed into a Jordan arc by adjunction of this point, i.e., by setting $\lambda(\beta) = Z$. A curvilinear half-interval of this kind will be called a *Jordan half-interval*, with final point Z.

Example 2. Let γ be the curvilinear half-interval with equation

$$z = t + i \sin \frac{1}{t} \qquad (\alpha \leqslant t < 0),$$

i.e., the graph of the function

$$y = \sin \frac{1}{x} \qquad (\alpha \leqslant x < 0).$$

Then the limit set E_γ consists of the whole segment $-1 \leqslant y \leqslant 1$ of the imaginary axis. In this case, $\gamma \cup E_\gamma$ is a continuum, but not a continuous curve.

THEOREM 2.5. *Let G be a domain with boundary Γ which is mapped conformally by a function $w = f(z)$ onto a domain D with boundary Δ. Suppose $\{z_n\}$ is a sequence of points in G all of whose limit points belong to Γ. Then all the limit points of the sequence $\{w_n\} = \{f(z_n)\}$ belong to Δ.*

[8] Here we use the term *Jordan arc* to denote a Jordan curve with distinct end points (cf. footnote 2, p. I.60).

Proof. The limit points of $\{w_n\}$ must belong to $\bar{D}$, since $w_n \in D$ $(n = 1, 2, \ldots)$. Suppose one of these limit points, say w_0, does not belong to Δ. Then $w_0 \in D$, and hence there exists a closed neighborhood $\overline{\mathcal{N}(w_0)} \subset D$. The inverse image of $\overline{\mathcal{N}(w_0)}$ is a compact set $F \subset G$ (see Theorem I.3.15). Since $\overline{\mathcal{N}(w_0)}$ contains infinitely many terms of the sequence $\{w_n\}$, F must contain infinitely many terms of the original sequence $\{z_n\}$. Therefore $\{z_n\}$ must have limit points in F and hence in G, contrary to hypothesis. This contradiction proves the theorem.

THEOREM 2.6. *Let G be a domain with boundary Γ which is mapped conformally by a function $w = f(z)$ onto a domain D with boundary Δ. Suppose $\gamma \subset G$ is a curvilinear half-interval whose limit set is contained in Γ. Then $f(\gamma) \subset D$ is a curvilinear half-interval whose limit set is contained in Δ.*

Proof. Suppose the equation of γ is

$$z = \lambda(t) \qquad (\alpha \leqslant t < \beta).$$

Then $f(\gamma)$ has the equation

$$w = f[\lambda(t)] \qquad (\alpha \leqslant t < \beta),$$

and is obviously a curvilinear half-interval, since both $\lambda(t)$ and $f(z)$ are one-to-one and continuous. Let $\{t_n\}$ be an arbitrary sequence of points in $[\alpha, \beta)$ such that $t_n \to \beta-$ as $n \to \infty$. By hypothesis, all the limit points of the sequence $\{z_n\} = \{\lambda(t_n)\}$ belong to Γ. But then, according to Theorem 2.5, all the limit points of the sequence $\{w_n\} = \{f[\lambda(t_n)]\}$ belong to Δ, as asserted.

In proving Theorems 2.5 and 2.6, the conformality of the mapping $w = f(z)$ was not actually used, but only the fact that $f(z)$ is a homeomorphism. We now consider deeper theorems, which exploit the conformality of $f(z)$:

THEOREM 2.7. *Let $\gamma: z = \lambda(t)$, $t \in [\alpha, \beta)$ be a curvilinear half-interval contained in the unit disk $|z| < 1$, whose limit set E_γ lies on the unit circle and consists of more than one point. Suppose the function $f(z)$ is analytic and bounded on the unit disk, and approaches a limit c along γ, i.e.,*

$$\lim_{t \to \beta-} f[\lambda(t)] = c.$$

Then $f(z)$ is identically equal to c.

Proof. Suppose the theorem is false, so that $f(z) - c \not\equiv 0$, i.e.,

$$f(z) - c = \delta z^k + \cdots \qquad (|z| < 1),$$

where $\delta \neq 0$. Then the function

$$\varphi(z) = \frac{f(z) - c}{z^k} = \delta + \cdots \qquad (|z| < 1)$$

takes the value $\varphi(0) = \delta \neq 0$ at the origin. Moreover, $\varphi(z)$ is bounded on the unit disk,[9] and approaches the limit 0 along γ. Consider two distinct points a and b belonging to E_γ, and let $\{a_n'\} = \{\lambda(t_n')\}$ and $\{b_n'\} = \{\lambda(\tau_n')\}$ be two sequences of points in γ which converge to a and b, respectively (so that both $t_n' \to \beta-$ and $\tau_n' \to \beta-$ as $n \to \infty$). It can be assumed that

$$t_1' < \tau_1' < \cdots < t_n' < \tau_n' < \cdots,$$

by replacing $\{t_n'\}$ and $\{\tau_n'\}$ by appropriate subsequences, if necessary. Let σ_n' denote the Jordan arc contained in γ with equation

$$z = \lambda(t) \qquad (t_n' \leqslant t \leqslant \tau_n').$$

Then for any r $(0 < r < 1)$, we can find an integer $N(r) > 0$ such that σ_n' is contained in the annulus $r < |z| < 1$ for all $n > N(r)$.

Next let σ' be the smaller of the two arcs of the unit circle with end points a and b, and on σ' choose an arc $\overset{\frown}{AB}$ with end points different from a and b, and of length π/m, where m is an integer greater than 1. Suppose we draw the diameters AOD and BOC, as shown in Figure 2.6. Then for sufficiently large $n > N'$, the point a_n' will lie in the domain g' (AA_1C_1CaA) and the point b_n' will lie in the domain d' (BB_1D_1DbB), while the arc σ_n' will lie in the annulus $r < |z| < 1$. It is easy to see that for every $n > N'$, there exists an arc $\sigma_n \subset \sigma_n'$ with one end point on the diameter AD and the other end point on the diameter BC, such that σ_n is contained entirely in either the domain AA_1B_1B or the domain CC_1D_1D, except for its end points. In fact, as t increases from t_n' to τ_n', the point $z = \lambda(t)$ describes

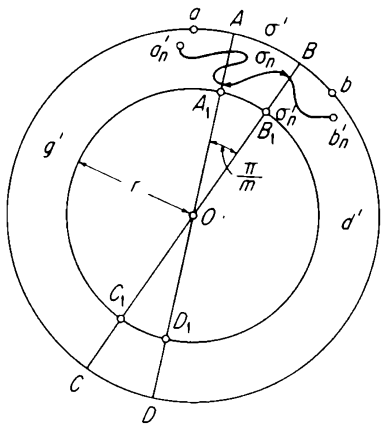


FIGURE 2.6

[9] If $0 \leqslant |z| \leqslant r < 1$, then, by the maximum modulus principle (Theorem I.17.5)

$$|\varphi(z)| \leqslant \frac{1}{r^k}\left(\max_{|z|=r} |f(z)| + |c|\right) \leqslant \frac{M + |c|}{r^k},$$

where

$$M = \sup_{|z|<1} |f(z)|.$$

Fixing z and making r approach 1, we have

$$|\varphi(z)| \leqslant M + |c| \qquad \text{if } |z| < 1.$$

the arc σ_n' continuously, going from the point a_n' to the point b_n', and moreover, a_n' and b_n' belong to domains g' and d' such that $\bar{g}' \cap \bar{d}' = 0$. Hence there must exist a number t_n in the interval (t_n', τ_n') such that $a_n = \lambda(t_n) \in \bar{g}'$ while $\lambda(t) \notin \bar{g}'$ for $t_n < t \leqslant \tau_n'$, and a number τ_n in the interval (t_n, τ_n') such that $b_n = \lambda(\tau_n) \in \bar{d}'$ while $\lambda(t) \notin \bar{d}'$ for $t_n \leqslant t < \tau_n$. The points a_n and b_n obviously lie on the boundaries of the domains g' and d', i.e., on the diameters AD and BC. In fact, a_n lies on the segment A_1A or C_1C, while b_n lies on the segment BB_1 or DD_1. Let σ_n denote the Jordan arc with equation

$$z = \lambda(t) \qquad (t_n \leqslant t \leqslant \tau_n).$$

Then $\sigma_n \subset \sigma_n'$ obviously has the required property, i.e., σ_n is contained in either AA_1B_1B or CC_1D_1D, except for its end points (a_n and b_n) which belong to the boundary of the appropriate domain. Since n takes infinitely many values, at least one of the domains AA_1B_1B and CC_1D_1D, say AA_1B_1B, contains infinitely many distinct arcs

$$\sigma_{n_1}, \sigma_{n_2}, \ldots, \sigma_{n_k}, \ldots \qquad (n_1 < n_2 < \cdots < n_k < \cdots). \tag{2.10}$$

We now make a change of variable

$$\zeta = e^{i\theta} z \tag{2.11}$$

such that the diameter BC goes into the real axis of the ζ-plane and the image of the domain AA_1B_1B lies in the first quadrant. Let $s_1, s_2, \ldots, s_k, \ldots$ denote the images of the arcs (2.10) under the mapping (2.11). The function $\varphi(z)$ is transformed by (2.11) into the function

$$\psi(\zeta) = \varphi(e^{-i\theta}\zeta),$$

which is analytic and bounded on the unit disk $|\zeta| < 1$, and satisfies the conditions $\psi(0) = \delta \neq 0$ and

$$\lim_{\substack{k \to \infty \\ \zeta \in s_k}} \psi(\zeta) = 0. \tag{2.12}$$

Writing

$$\mu_k = \max_{\zeta \in s_k} |\psi(\zeta)|,$$

we can express (2.12) in the form

$$\lim_{k \to \infty} \mu_k = 0. \tag{2.13}$$

Suppose $\psi(\zeta)$ has the Taylor series

$$\psi(\zeta) = \delta + \delta_1\zeta + \cdots + \delta_n\zeta^n + \cdots \qquad (|\zeta| < 1).$$

Then, introducing the function

$$\psi^*(\zeta) = \bar{\delta} + \bar{\delta}_1\zeta + \cdots + \bar{\delta}_n\zeta^n + \cdots \qquad (|\zeta| < 1),$$

we see that $\psi(\zeta)$ and $\psi^*(\zeta)$ takes conjugate complex values at points which are symmetric with respect to the real axis, i.e.,

$$\psi^*(\zeta) = \overline{\psi(\bar{\zeta})}.$$

Obviously, $\psi^*(\zeta)$ is analytic and bounded on the unit disk, and satisfies the conditions $\psi^*(0) = \overline{\psi(0)} \neq 0$ and

$$\sup_{|\zeta|<1} |\psi(\zeta)| = \sup_{|\zeta|<1} |\psi^*(\zeta)| = M < \infty.$$

Moreover, if s_k^* denotes the arc symmetric to s_k with respect to the real axis, then

$$\max_{\zeta \in s_k} |\psi(\zeta)| = \max_{\zeta \in s_k^*} |\psi^*(\zeta)| = \mu_k.$$

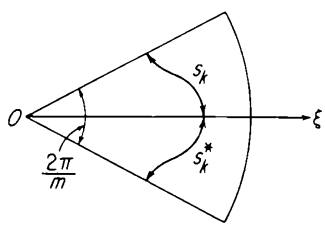


FIGURE 2.7

The arcs s_k and s_k^* share a common point lying on the real axis, and together they form a Jordan arc S_k, which is symmetric with respect to the real axis and joins the sides of an angle of $2\pi/m$ radians bisected by the real axis (see Figure 2.7, where $\xi = \operatorname{Re} \zeta$). The analytic function

$$\chi(\zeta) = \psi(\zeta)\psi^*(\zeta)$$

satisfies the conditions

$$|\chi(\zeta)| \leqslant M^2 \qquad (|\zeta| < 1), \qquad \chi(0) = \psi(0)\psi^*(0) = |\delta|^2 \neq 0$$

and

$$\max_{\zeta \in S_k} |\chi(\zeta)| \leqslant M\mu_k.$$

Subjecting the arc S_k to the rotations

$$\zeta' = e^{2\pi i j/m} \qquad (j = 1, 2, \ldots, m-1),$$

we obtain new arcs $S_k^{(1)}, \ldots, S_k^{(m-1)}$, which together with S_k form a closed Jordan curve Γ_k contained in the unit disk, where the interior of Γ_k obviously contains the origin. If ζ is an arbitrary point of Γ_k, then all m points

$$\zeta, e^{2\pi i/m}\zeta, \ldots, e^{2\pi(m-1)i/m}\zeta$$

lie on Γ_k, and at least one of them belongs to S_k. Therefore at the point $\zeta \in \Gamma_k$, at least one of the m numbers

$$\chi(\zeta), \chi(e^{2\pi i/m}\zeta), \ldots, \chi(e^{2\pi(m-1)i/m}\zeta)$$

has absolute value no greater than $M\mu_k$, while the remaining numbers have absolute values no greater than M^2. It follows that the product

$$F(\zeta) = \chi(\zeta)\chi(e^{2\pi i/m}\zeta)\cdots\chi(e^{2\pi i(m-1)/m}\zeta)$$

satisfies the inequality

$$|F(\zeta)| \leqslant M\mu_k M^{2(m-1)}$$

at every point $\zeta \in \Gamma_k$. Since the function $F(\zeta)$ is analytic, we also have

$$|F(0)| = |\delta|^{2m} \leqslant \mu_k M^{2m-1}.$$

But then, using (2.13) to take the limit as $k \to \infty$, we find that

$$\delta = 0,$$

contrary to hypothesis. This contradiction proves the theorem.

COROLLARY. *Let $\{\gamma_n\}$ be a sequence of Jordan arcs contained in the unit disk $|z| < 1$ (with the possible exception of initial and final points) such that every sequence $\{z_n\}$, $z_n \in \gamma_n$ has two or more limit points, all lying on the unit circle. Suppose the function $f(z)$ is analytic and bounded on the unit disk, and satisfies the condition*

$$\lim_{\substack{n \to \infty \\ z_n \in \gamma_n}} f(z_n) = c.$$

Then $f(z)$ is identically equal to c.

Proof. Note that on the sequence of arcs $\{\gamma_n\}$, we can construct a sequence of arcs $\{\sigma_n\}$ just like those used to prove Theorem 2.7.

THEOREM 2.8. *Let g be a domain bounded by a closed Jordan curve γ such that $g \cup \gamma$ is contained in the unit disk $|z| < 1$, except for a single boundary point ζ lying on the unit circle. Let*

$$\gamma_1 : z = \lambda_1(t) \quad (\alpha_1 \leqslant t < \beta_1), \qquad \gamma_2 : z_2 = \lambda_2(t) \quad (\alpha_2 \leqslant t < \beta_2)$$

be two Jordan half-intervals with the common final point ζ, such that $\gamma = \gamma_1 \cup \gamma_2 \cup \{\zeta\}$. Suppose the function $f(z)$ is analytic and bounded on the unit disk, and approaches definite limits c_1 and c_2 along γ_1 and γ_2, i.e.,

$$\lim_{t \to \beta_1 -} f[\lambda_1(t)] = c_1, \qquad \lim_{t \to \beta_2 -} f[\lambda_2(t)] = c_2. \tag{2.14}$$

Then $c_1 = c_2 = c$, and moreover

$$\lim_{\substack{z \to \zeta \\ z \in g}} f(z) = c.$$

Proof. First we consider the case where $c_1 = c_2 = 0$ and prove the second part of the theorem, i.e., that

$$\lim_{\substack{z \to \zeta \\ z \in g}} f(z) = 0.$$

Obviously, there is no need to consider points on the boundary of the domain g, since for such points $f(z) \to 0$ by hypothesis. Thus let z_0 be

an interior point of g, and draw the chord δ_0 through z_0 perpendicular to the radius $O\zeta$, as shown in Figure 2.8. This chord divides g into subdomains, some of which lie on the same side of δ_0 as the point ζ. Among these latter subdomains, there is a unique subdomain g_0 with the point z_0 on its boundary. The point ζ may not be a boundary point of g_0, in which case the closed domain $\bar{g}_0$ is contained in the unit disk (Figure 2.8 shows the other case, where ζ is a boundary point of g_0). The boundary of g_0 consists of the parts of the Jordan half-intervals γ_1 and γ_2 which lie on the same side of δ_0 as ζ, together with segments of δ_0 (which may degenerate into points). We make the further assumption that $|f(z)| < \varepsilon$ at the boundary points of g_0 belonging to $\gamma_1 \cup \gamma_2$, where ε is a preassigned (but arbitrary) positive number. This can always be achieved by choosing z_0 sufficiently close to ζ, since, by hypothesis, $f(z) \to 0$ along γ_1 and γ_2.

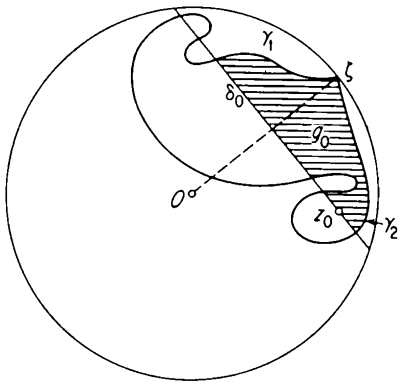


FIGURE 2.8

We now make a change of variable

$$z' = e^{i\theta}z + a \tag{2.15}$$

(a rotation and a shift), choosing θ and a in such a way that the chord δ_0 goes into a segment of the real axis, while ζ goes into a point ζ' of the positive imaginary axis, as shown in Figure 2.9. Then the domain g_0 goes into a domain g_0' lying in the upper half-plane, the point $z = 0$ goes into the point a on the negative imaginary axis, the unit disk $|z| < 1$ goes into the disk $|z' - a| < 1$, γ_1 and γ_2 go into new Jordan half-intervals γ_1' and γ_2', and finally, z_0 goes into a point z_0' on the real axis, where z_0' is a boundary point of g_0'. The function $f(z)$ is transformed by (2.15) into the function

$$\varphi(z') = f[(z' - a)e^{-i\theta}],$$

which is analytic and bounded on the disk $|z' - a| < 1$, and satisfies the inequality $|\varphi(z')| < \varepsilon$ at the boundary points of g_0'

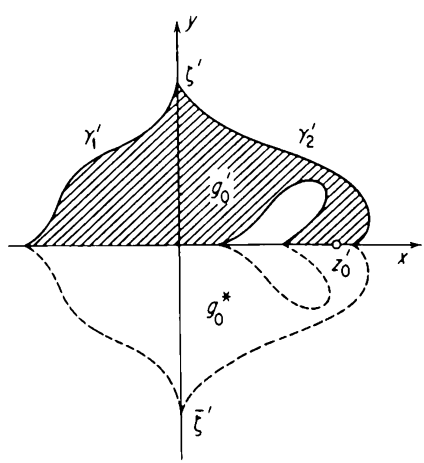


FIGURE 2.9

belonging to $\gamma_1' \cup \gamma_2'$. Suppose $\varphi(z')$ has the Taylor series

$$\varphi(z') = \sum_{n=0}^{\infty} c_n(z' - a)^n.$$

Then, introducing the function

$$\varphi^*(z') = \sum_{n=0}^{\infty} \bar{c}_n(z' - \bar{a})^n,$$

we see that $\varphi(z')$ and $\varphi^*(z')$ take conjugate complex values at points which are symmetric with respect to the real axis, i.e.,

$$\varphi^*(z') = \overline{\varphi(\bar{z}')}.$$

Obviously, the function $\varphi^*(z')$ is analytic and bounded on the disk $|z' - \bar{a}| < 1$. Moreover, if g_0^* is the domain symmetric to g_0' with respect to the real axis, then $|\varphi^*(z')| < \varepsilon$ at the boundary points of g_0^* which lie below the real axis. Let G_0 be the domain consisting of the domains g_0' and g_0^*, together with all segments of the real axis consisting of common boundary points of g_0' and g_0^*. Then z_0' is an interior point of G_0. Writing

$$\sup_{|z'-a|<1} |\varphi(z')| = \sup_{|z|<1} |f(z)| = M < \infty,$$

we have

$$\sup_{|z'-\bar{a}|<1} |\varphi^*(z')| = M.$$

It follows that the function $\varphi(z')\varphi^*(z')$, which is analytic on G_0, satisfies the inequality

$$|\varphi(z')\varphi^*(z')| \leqslant M^2$$

for all $z' \in G_0$.

Actually, as we now show, $\varphi(z')\varphi^*(z')$ satisfies the stronger inequality

$$|\varphi(z')\varphi^*(z')| \leqslant M\varepsilon$$

for all $z' \in G_0$. There are two cases to consider. First suppose ζ does not belong to the boundary of the domain g_0. Then ζ' does not belong to the boundary of g_0', so that $\bar{g}_0'$ is contained in the disk $|z' - a| < 1$ and $\bar{g}_0^*$ is contained in the disk $|z' - \bar{a}| < 1$. Then the function $\varphi(z')\varphi^*(z')$ is analytic and hence continuous on the closed domain $\bar{G}_0$. Therefore, since

$$|\varphi(z')\varphi^*(z')| \leqslant M\varepsilon$$

on the boundary of G_0, it follows from the maximum modulus principle that this inequality holds for all $z' \in G_0$, in particular for z_0':

$$|\varphi(z_0')\varphi^*(z_0')| \leqslant M\varepsilon. \tag{2.16}$$

Noting that

$$\varphi^*(z_0') = \overline{\varphi(\bar{z}_0')} = \overline{\varphi(z_0')}$$

($\bar{z}_0' = z_0'$, since z_0' is real), we can write (2.16) as

$$|\varphi(z_0')|^2 = |f(z_0)|^2 \leqslant M\varepsilon. \tag{2.17}$$

Next suppose ζ belongs to the boundary of G_0. Then the boundary points ζ' and $\bar{\zeta}'$ of the domain G_0 are not interior points of the disks $|z' - a| < 1$ and $|z' - \bar{a}| < 1$, and we can no longer assert that $\varphi(z')\varphi^*(z')$ is continuous on the closed domain $\bar{G}_0$. To reduce this case to the previous case of a function continuous on $\bar{G}_0$, we construct the new function

$$\psi(z') = (z' - \zeta')^\mu (z' - \bar{\zeta}')^\mu \varphi(z')\varphi^*(z'),$$

where μ is an arbitrary positive number, and by $(z' - \zeta')^\mu$ and $(z' - \bar{\zeta}')^\mu$ we mean arbitrary single-valued analytic branches of the corresponding multiple-valued functions. Then the function $\psi(z')$ is single-valued and analytic on G_0, and bounded on G_0 as well, since

$$|\psi(z')| < 2^\mu 2^\mu M^2 = 4^\mu M^2 \qquad (z' \in G_0).$$

Moreover, $\psi(z')$ is continuous on the closed domain $\bar{G}_0$ if we set

$$\psi(\zeta') = \lim_{z' \to \zeta'} (z' - \zeta')^\mu (z' - \bar{\zeta}')^\mu \varphi(z')\varphi^*(z') = 0,$$

and similarly $\psi(\bar{\zeta}') = 0$. Therefore, since

$$|\psi(z')| < 2^\mu 2^\mu M\varepsilon = 4^\mu M\varepsilon$$

on the boundary of G_0, it follows from the maximum modulus principle that this inequality holds for all $z' \in G_0$, in particular for z_0':

$$|\psi(z_0')| = |\varphi(z_0')|^2 |z_0 - \zeta'|^{2\mu} < 4^\mu M\varepsilon. \tag{2.18}$$

Taking the limit as $\mu \to 0$ in (2.18), we obtain

$$|\varphi(z_0')|^2 = |f(z_0)|^2 \leqslant M\varepsilon,$$

which is the same as (2.17). Thus, in any case

$$|f(z_0)|^2 \leqslant M\varepsilon$$

for arbitrary $\varepsilon > 0$ and $z_0 \in g$ (provided z_0 is sufficiently close to ζ). It follows that

$$\lim_{\substack{z_0 \to \zeta \\ z_0 \in g}} f(z_0) = 0,$$

and the second half of the theorem is proved.[10]

We must still prove the first part of the theorem, which asserts that $c_1 = c_2$ is implied by (2.14), or equivalently by

$$\lim_{\substack{z \to \zeta \\ z \in \gamma_1}} f(z) = c_1, \qquad \lim_{\substack{z \to \zeta \\ z \in \gamma_2}} f(z) = c_2.$$

[10] If $c_1 = c_2 = c \neq 0$, consider the function $f(z) - c$ instead of $f(z)$.

Consider the function

$$F(z) = [f(z) - c_1][f(z) - c_2].$$

This function satisfies the conditions of the theorem, with (2.14) replaced by

$$\lim_{\substack{z\to\zeta \\ z\to\gamma_1}} F(z) = \lim_{\substack{z\to\zeta \\ z\in\gamma_2}} F(z) = 0.$$

Suppose $c_1 \neq c_2$, and let $|c_1 - c_2| = \rho > 0$. Moreover, let the neighborhood $\mathcal{N}(\zeta)$ be so small that

$$|F(z)| = |f(z) - c_1|\,|f(z) - c_2| < \frac{\rho^2}{4} \qquad (z \in \bar{g}) \tag{2.19}$$

and

$$|f(z) - c_1| < \frac{\rho}{2} \qquad (z \in \gamma_1), \quad |f(z) - c_2| < \frac{\rho}{2} \qquad (z \in \gamma_2) \tag{2.20}$$

(this is possible by the part of the theorem already proved). Then, among the domains whose union equals $\mathcal{N}(\zeta) \cap g$, there is a domain g_0 with the point ζ on its boundary. Let ζ_1 and ζ_2 be two boundary points of g_0 which belong to $\mathcal{N}(\zeta) \cap \gamma_1$ and $\mathcal{N}(\zeta) \cap \gamma_2$, respectively. Since the first of the inequalities (2.20) is satisfied by $z = \zeta_1$ and the second by $z = \zeta_2$, then, by the continuity of $f(z)$, we can find two points z_1 and z_2 in g_0 lying so close to ζ_1 and ζ_2, respectively, that

$$|f(z_1) - c_1| < \frac{\rho}{2}, \qquad |f(z_2) - c_2| < \frac{\rho}{2},$$

which, in particular, implies

$$|f(z_2) - c_1| \geqslant |c_2 - c_1| - |f(z_2) - c_2| > \frac{\rho}{2}.$$

Now let $\gamma \subset g_0$ be a Jordan curve joining z_1 and z_2 (cf. Theorem I.4.12), so that $f(\gamma)$ is a Jordan curve joining the points $f(z_1)$ and $f(z_2)$. Since the disk $|w - c_1| < \rho/2$ contains $f(z_1)$ but not $f(z_2)$, the curve γ must contain a point z_0 such that

$$|f(z_0) - c_1| = \frac{\rho}{2}.$$

Then

$$|f(z_0) - c_2| \geqslant |c_2 - c_1| - |f(z_0) - c_1| = \frac{\rho}{2},$$

i.e., we have found a point $z_0 \in g_0$ such that

$$|F(z_0)| = |f(z_0) - c_1|\,|f(z_0) - c_2| \geqslant \frac{\rho^2}{4},$$

contrary to (2.19). This contradiction shows that $c_1 = c_2$, thereby completing the proof of the theorem.

6. Accessible Boundary Points

Given a domain G with boundary Γ, consider the set E of all Jordan half-intervals $\gamma \subset G$ whose final points belong to Γ. In general, Γ will contain points which cannot be final points of Jordan half-intervals in E. For example, suppose we draw infinitely many line segments

$$x = \frac{1}{2n}, \quad 0 \leqslant y \leqslant \frac{3}{4} \qquad (n = 1, 2, \ldots)$$

and

$$x = \frac{1}{2n+1}, \quad \frac{1}{4} \leqslant y \leqslant 1 \qquad (n = 1, 2, \ldots)$$

in the closed square $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$, and then take G to be the domain whose boundary Γ consists of these segments together with the four sides of the square (see Figure 2.10). Then, obviously, no boundary point belonging to the imaginary axis can be the final point of a Jordan half-interval contained in G. However, in any case, it is easy to see that the set of points which can serve as final points of Jordan half-intervals $\gamma \in E$ is everywhere dense in Γ, i.e., every $\zeta \in \Gamma$ is a limit point of such points. In fact, let $\mathcal{N}(\zeta)$ be any neighborhood of ζ and z_0 any point in $\mathcal{N}(\zeta) \cap G$, and draw the line segment δ joining ζ and z_0. Then the point of $\delta \cap \Gamma$ closest to z_0 (which may be ζ itself) is the final point of a Jordan half-interval in E, namely, a rectilinear half-interval with initial point z_0.

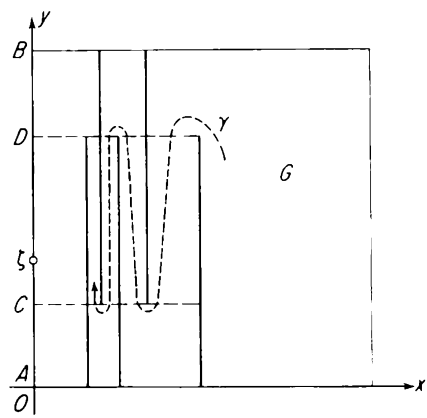


FIGURE 2.10

DEFINITION. *Every Jordan half-interval $\gamma \in E$ is said to determine an accessible point of the boundary Γ (relative to the domain G). Moreover, two half-intervals $\gamma_1 \in E$ and $\gamma_2 \in E$ are said to determine the same accessible boundary point if the following two conditions are met:*

1. *γ_1 and γ_2 have the same final point ζ;*
2. *There exists a half-interval $\gamma_3 \in E$ with final point ζ such that γ_3 shares points with both γ_1 and γ_2 in every neighborhood $\mathcal{N}(\zeta)$.*

Thus, for example, Figure 2.11 shows a domain G whose boundary consists of a circle together with one of its radii. Here the half-intervals γ_1 and γ_2 determine the same accessible boundary point, but γ_1 and γ' (or γ_2 and γ') determine different accessible boundary points.

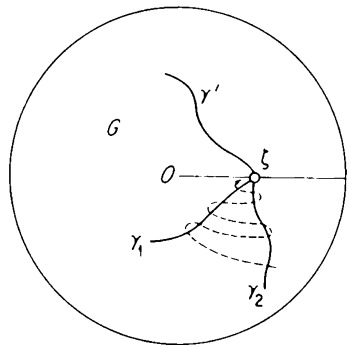


FIGURE 2.11

The final point of a Jordan half-interval γ will be regarded as *representing* an accessible boundary point. It should be kept in mind that in this sense, the same point ζ may represent two or more different accessible boundary points. For example, consider the domain G shown schematically in Figure 2.12 consisting of the "half-disk"

$$0 < \arg z < \pi, \qquad 0 < |z| < 1,$$

after eliminating all rays of the form

$$\arg z = \frac{p}{q}\pi, \qquad 0 < |z| \leqslant \frac{1}{q},$$

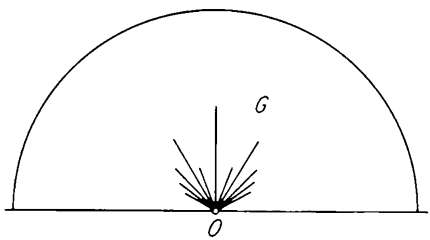


FIGURE 2.12

where p and q $(p < q)$ are relatively prime positive integers. For this domain, the origin O represents an uncountable set of different accessible points (actually, a set with the power of the continuum). In fact, if α $(0 < \alpha < 1)$ is irrational, then every rectilinear half-interval

$$\arg z = \alpha\pi, \qquad 0 < |z| \leqslant \frac{1}{2}$$

determines an accessible point represented by O, and distinct values of α obviously correspond to distinct accessible points.

We now prove some theorems which show how Jordan half-intervals behave under conformal mappings:

THEOREM 2.9. *Suppose a bounded simply connected domain G, with boundary Γ, is mapped conformally by a function $w = f(z)$ onto the unit disk $K: |w| < 1$.*[11] *Then the set of all Jordan half-intervals contained in G determining the same accessible point of Γ is transformed into a set of Jordan half-intervals contained in K, all with the same final point σ on the unit circle $\Sigma: |w| = 1$.*

[11] The domain G must be simply connected, being the homeomorphic image of a disk (see Theorem I.6.4). The boundedness of G is needed to apply Theorem 2.7.

Proof. Let E be the set of all Jordan half-intervals contained in G with final points on Γ. If γ is a half-interval in E whose final point ζ determines an accessible boundary point, then, according to Theorem 2.6, $\delta = f(\gamma)$ is a curvilinear half-interval contained in K, and all the limit points of δ lie on Σ. Suppose δ is not a Jordan half-interval, i.e., suppose δ has at least two limit points on Σ. Then, applying Theorem 2.7 to δ and the inverse function $z = f^{-1}(w)$, after noting that[12]

$$\lim_{\delta} f^{-1}(w) = \zeta, \tag{2.21}$$

we find that

$$f^{-1}(w) \equiv \zeta,$$

contrary to hypothesis. Therefore δ is a Jordan half-interval, whose final point σ lies on Σ.

Next suppose $\gamma_1 \in E$ and $\gamma_2 \in E$ determine the same accessible point of Γ, so that γ_1 and γ_2 have the same final point ζ, and let σ_1 and σ_2 denote the final points of the half-intervals $\delta_1 = f(\gamma_1)$ and $\delta_2 = f(\gamma_2)$. By the definition of an accessible point, there exists a half-interval $\gamma_3 \in E$ such that γ_3 intersects γ_1 and γ_2 in every neighborhood of ζ. Clearly, $\delta_3 = f(\gamma_3)$ is a Jordan half-interval which has an infinite set of points in common with both δ_1 and δ_2. Moreover, the points σ_1 and σ_2 are limit points of this set (see Theorem 2.5), and hence belong to the limit set of the Jordan half-interval δ_3. But this is possible only if $\sigma_1 = \sigma_2 = \sigma$. The proof is now complete.

Theorem 2.9 establishes a correspondence between accessible points of the boundary Γ and certain points of the circle Σ, which are called the *images* of the accessible points under the mapping $w = f(z)$. Next we show that this correspondence is one-to-one:

THEOREM 2.10. *With the same notation as in Theorem* 2.9, *distinct accessible points of* Γ *have distinct images under the mapping* $w = f(z)$.

Proof. Suppose the half-intervals $\gamma_1 \in E$ and $\gamma_2 \in E$ with final points ζ_1 and ζ_2 determine distinct accessible points of the boundary Γ. Let σ_1 and σ_2 be the final points of the half-intervals $\delta_1 = f(\gamma_1)$ and $\delta_2 = f(\gamma_2)$. Suppose, contrary to the assertion of the theorem, that $\sigma_1 = \sigma_2 = \sigma$. Then two cases are possible. In the first case, δ_1 and δ_2 have infinitely many common points in every neighborhood of σ. Then γ_1 and γ_2 must have an infinite set of points in common, and this set has all its limit points on Γ (see Theorem 2.5). But then γ_1 and γ_2 have the same final point, and hence determine the same accessible point of Γ (why?). This contradiction establishes the theorem for the first case.

---

[12] By the left-hand side of (2.21) is meant the limit of z as w approaches Σ along δ.

In the second case, δ_1 and δ_2 have only finitely many common points in some neighborhood of the point σ. Then, by deleting certain initial parts of the half-intervals δ_1 and δ_2, we obtain two disjoint Jordan half-intervals δ_1' and δ_2' with the same final point σ. The inverse images of δ_1' and δ_2' in the domain G are Jordan half-intervals γ_1' and γ_2', obtained by deleting certain initial parts from γ_1 and γ_2. Obviously, γ_1' and γ_2' determine the same accessible points of the boundary Γ as γ_1 and γ_2. With the intention of eventually applying Theorem 2.8, we join any two points $w_1' \in \delta_1'$ and $w_2' \in \delta_2'$ by a Jordan arc $\delta_0' \subset K$. Traversing δ_0' from w_1' to w_2', we delete both the initial part of δ_0' going from w_1' to the last point w_1 in which δ_0' intersects δ_1', and the final part of δ_0' going from the first point w_2 in which δ_0' intersects δ_2' to the point w_2'. In this way, we obtain a new Jordan arc $\tilde{\delta}_0$ joining $w_1 \in \delta_1'$ and $w_2 \in \delta_2'$, which has no points in common with δ_1' and δ_2' other than w_1 and w_2. Finally, deleting the initial parts of δ_1' and δ_2' up to the points w_1 and w_2, respectively, we obtain new Jordan half-intervals $\tilde{\delta}_1$ and $\tilde{\delta}_2$, which together with their common point σ and the arc $\tilde{\delta}_0$ form a closed Jordan curve, bounding some subdomain $g \subset K$. By construction, the inverse images of the Jordan half-intervals $\tilde{\delta}_1$ and $\tilde{\delta}_0 + \tilde{\delta}_2$ (concerning the meaning of the plus sign, see footnote 12, p. 1.156) in the domain G are also Jordan half-intervals $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ with final points ζ_1 and ζ_2, determining the same accessible points as the original half-intervals γ_1 and γ_2. For the function $z = f^{-1}(w)$, we have

$$\lim_{\tilde{\delta}_1} f^{-1}(w) = \zeta_1, \qquad \lim_{\tilde{\delta}_0 + \tilde{\delta}_2} f^{-1}(w) = \zeta_2.$$

Therefore it follows from Theorem 2.8 that $\zeta_1 = \zeta_2 = \zeta$, i.e., $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ have the same final point ζ.

Next, in the closed domain $\bar{g}$, we construct a Jordan half-interval $\tilde{\delta}_3$ with final point σ, which shares points with both $\tilde{\delta}_1$ and $\tilde{\delta}_0 + \tilde{\delta}_2$ in every neighborhood $\mathcal{N}(\sigma)$. Such a half-interval can be made up of arcs from an infinite family of concentric circles of radii $\varepsilon_1, \ldots, \varepsilon_n, \ldots$ with the point σ as center, where $\varepsilon_{n+1} < \varepsilon_n$ $(n = 1, 2, \ldots)$ and $\varepsilon_n \to 0$ as $n \to \infty$, together with alternate arcs of $\tilde{\delta}_1$ and $\tilde{\delta}_0 + \tilde{\delta}_2$ joining consecutive circles. This construction is shown schematically in Figure 2.13, where $\tilde{\delta}_3$ is the dotted

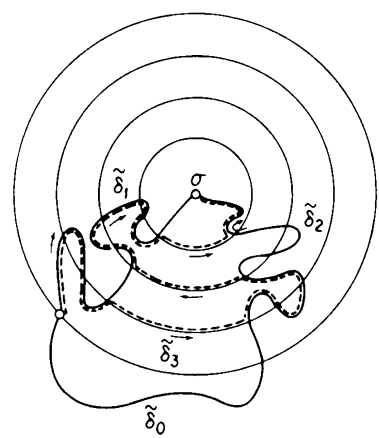


FIGURE 2.13

curve. The function $z = f^{-1}(w)$ maps $\tilde{\delta}_3$ onto a curvilinear half-interval $\tilde{\gamma}_3 \subset G$, with its limit set contained in Γ (see Theorem 2.6). Moreover, according to Theorem 2.8, a point z describing $\tilde{\gamma}_3$ must converge to the point $\zeta \in \Gamma$. Therefore $\tilde{\gamma}_3$ is a Jordan half-interval with final point ζ. Since $\tilde{\gamma}_3$ has infinitely many points in common with both $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ in any neighborhood of the point ζ, it follows that $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ determine the same accessible point of Γ, contrary to hypothesis. This completes the proof.

To recapitulate, under conformal mapping of the domain G onto the disk $|w| < 1$, the set of accessible boundary points of G is transformed into a set e of points of the circle $\Sigma: |w| = 1$, and the correspondence between the accessible points and e is one-to-one. In particular, if G is the domain shown in Figure 2.11, the two accessible points represented by the point ζ must go into two distinct points of the unit circle, while if G is the domain shown in Figure 2.12, the accessible points represented by the point O must go into infinitely many distinct points of Σ (in fact, into a set with the power of the continuum).

THEOREM 2.11. *With the same notation as in Theorem* 2.9, *the set e of images of accessible points of* Γ *under the mapping* $w = f(z)$ *is dense in the unit circle* Σ.

Proof. Let $\{w_n\}$ be a sequence of points of the unit disk converging to a point $w_0 \in \Sigma$, and let $\{z_n\} = \{f^{-1}(w_n)\}$ be the corresponding sequence of points of the domain G. According to Theorem 2.5, all the limit

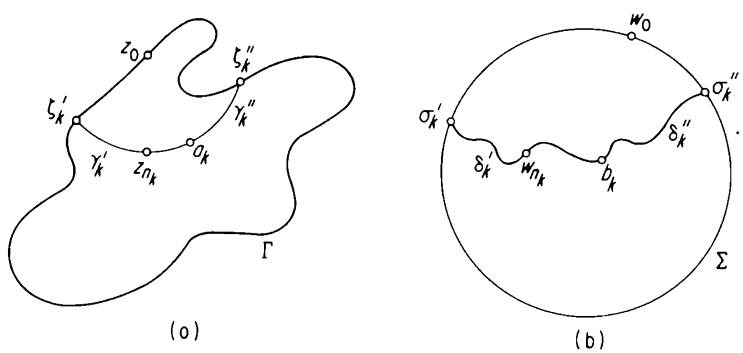


FIGURE 2.14

points of the sequence $\{z_n\}$ belong to Γ. From $\{z_n\}$ we choose a subsequence $\{z_{n_k}\}$ converging to a point $z_0 \in \Gamma$, subject to the requirement that the distances $|z_{n_k} - z_0|$ decrease as $n_k \to \infty$. Then the sequence $\{w_{n_k}\} = \{f(z_{n_k})\}$ converges to w_0, as before. Through each point z_{n_k} we draw an arc of the circle with center at z_0, extending the arc up to the points ζ_k' and ζ_k'' in which it first intersects Γ; as shown in Figure 2.14(a),

this arc consists of a pair of Jordan half-intervals γ_k' and γ_k'', with final points ζ_k' and ζ_k'', and common initial point z_{n_k}. According to Theorem 2.9, the images δ_k' and δ_k'' of these half-intervals are also Jordan half-intervals, with final points σ_k' and σ_k'' on the unit circle Σ and common initial point w_{n_k} [see Figure 2.14(b)]. Obviously, σ_k' and σ_k'' are the images of the distinct accessible points determined by γ_k' and γ_k''.

Next we note that, by construction, any sequence $\{a_k\}$, $a_k \in \gamma_k' \cup \gamma_k''$ has the unique limit point $z_0 \in \Gamma$. Therefore, by Theorem 2.5, the limit points of any sequence $\{b_k\}$, $b_k \in \delta_k' \cup \delta_k''$ belong to Σ. Moreover, by the corollary to Theorem 2.7, no such sequence $\{b_k\}$ can have more than one limit point, since otherwise it would follow from

$$\lim_{k \to \infty} f^{-1}(b_k) = z_0$$

that $f^{-1}(w) \equiv z_0$, contrary to hypothesis. But the sequence $\{w_{n_k}\}$, $w_{n_k} \in \delta_k' \cup \delta_k''$ converges to w_0, and hence the same is true of an arbitrary sequence $\{b_k\}$, $b_k \in \delta_k' \cup \delta_k''$. Therefore the sequences $\{\sigma_k'\}$ and $\{\sigma_k''\}$ of final points of the half-intervals δ_k' and δ_k'' must also converge to w_0, i.e., the arbitrary point $w_0 \in \Sigma$ is a limit point of images of accessible points of the boundary Γ. In other words, e is dense in Σ, as asserted.

7. Prime Ends

In this section we continue our study of the boundary structure of simply connected domains, eventually arriving at the key notion of a *prime end* (synonymously, *boundary element*) due to Carathéodory.

THEOREM 2.12. *Let G be a bounded simply connected domain with boundary Γ, and let $\gamma_1 \subset G$ and $\gamma_2 \subset G$ be two Jordan half-intervals determining two distinct accessible points of Γ. Then the set $\gamma_1 \cap \gamma_2$ has no limit points in Γ.*

Proof. Suppose the opposite, i.e., suppose $\gamma_1 \cap \gamma_2$ has limit points in Γ. Then γ_1 and γ_2 must have the same final point ζ (why?). Moreover, since every neighborhood $\mathcal{N}(\zeta)$ contains points of both γ_1 and γ_2, by choosing arcs alternately on γ_1 and γ_2, we can construct another Jordan half-interval $\gamma_3 \subset G$ with final point ζ which shares points with both γ_1 and γ_2 in every $\mathcal{N}(\zeta)$. But this contradicts the hypothesis that γ_1 and γ_2 determine distinct accessible points, thereby proving the theorem.

Remark. Obviously, by deleting certain initial parts of γ_1 and γ_2 (if necessary), we can find new Jordan half-intervals γ_1' and γ_2' such that $\gamma_1' \cap \gamma_2' = 0$. Thus there is no loss of generality in assuming from the outset that distinct accessible points are determined by disjoint Jordan half-intervals γ_1 and γ_2 (however, note that γ_1 and γ_2 may still have the same final point.)

Now let $\gamma_1 \subset G$ and $\gamma_2 \subset G$ be two disjoint Jordan half-intervals (with final points ζ_1 and ζ_2) determining two distinct accessible points of Γ, and draw a Jordan arc $\gamma \subset G$ going from the initial point z_1 of γ_1 to the initial point z_2 of γ_2. Let z_1' be the last point in which γ intersects γ_1, and delete the parts of γ_1 and γ preceding z_1'. Similarly, let z_2' be the first point in which γ intersects γ_2, and delete the part of γ_2 preceding z_2' and the part of γ following z_2'. As a result, we obtain two new Jordan half-intervals γ_1' and γ_2' determining the same accessible boundary points as γ_1 and γ_2, together with a Jordan arc γ' joining the initial point z_1' of γ_1' to the initial point z_2' of γ_2' [see Figure 2.15(a), which illustrates the case where γ_1' and γ_2' have distinct final points ζ_1 and ζ_2]. The set $\tilde{\gamma} = \gamma_1' \cup \gamma' \cup \gamma_2'$ is an open Jordan interval contained in G,[13] while the set $\tilde{\gamma} \cup \{\zeta_1\} \cup \{\zeta_2\}$ is a Jordan curve with end points on Γ and all other points in G.

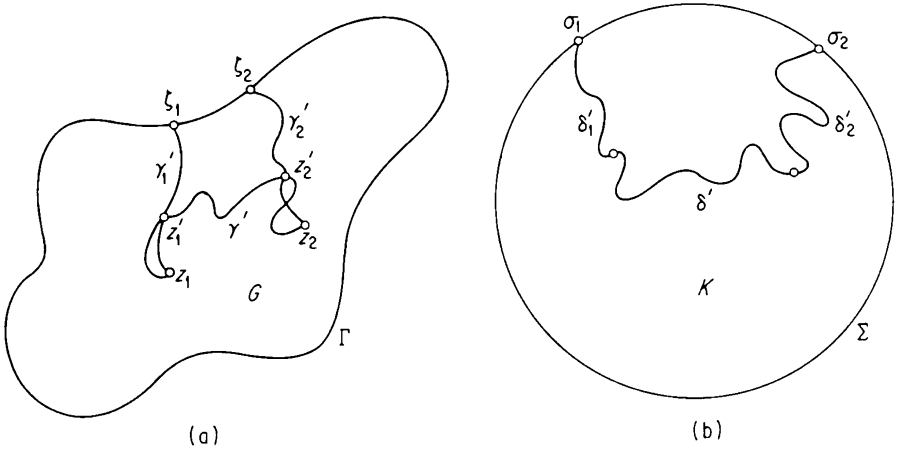


FIGURE 2.15

Under a conformal mapping $w = f(z)$ carrying G into the unit disk K, γ_1' and γ_2' go into Jordan half-intervals $\delta_1' \subset K$ and $\delta_2' \subset K$ with distinct final points σ_1 and σ_2 on the unit circle Σ (see Theorem 2.10), and γ' goes into a Jordan arc $\delta' \subset K$ [see Figure 2.15(b)]. The set $\tilde{\delta} = f(\tilde{\gamma}) = \delta_1' \cup \delta' \cup \delta_2'$ is an open Jordan interval contained in K, while the set $\tilde{\delta} \cup \{\sigma_1\} \cup \{\sigma_2\}$ is a Jordan curve with distinct end points on Σ and all other points in K. The end points σ_1 and σ_2 divide Σ into two closed arcs Σ_1 and Σ_2,[14] said to be *complementary* in the sense that $\Sigma = \Sigma_1 \cup \Sigma_2$. Together with $\tilde{\delta}$, the arcs Σ_1 and Σ_2 form two closed Jordan curves with interiors Δ_1 and Δ_2 contained

---

[13] I.e., $\tilde{\gamma}$ is the continuous one-to-one image of an open interval $\alpha < t < \beta$, and the limit set of $\tilde{\gamma}$ contains no more than two points. Here we are not concerned with orientation of curves (the set $\gamma_1' \cup \gamma' \cup \gamma_2'$ corresponds to two oriented curves $-\gamma_1' + \gamma' + \gamma_2'$ and $-\gamma_2' - \gamma' + \gamma_1'$).

[14] A circular arc is said to be *closed* if it contains its end points.

in K [see Figure 2.16(b)]. The two domains Δ_1 and Δ_2 (into which $\tilde{\delta}$ divides K in the sense that $K = \Delta_1 \cup \tilde{\delta} \cup \Delta_2$) are said to be *adjacent* to Σ_1 and Σ_2, respectively (for a geometrically obvious reason).

Returning to the z-plane, we find that $\tilde{\gamma}$ divides G into two simply connected domains $g_1 = f^{-1}(\Delta_1)$ and $g_2 = f^{-1}(\Delta_2)$, i.e., $G = g_1 \cup \tilde{\gamma} \cup g_2$ [see Figure 2.16(a)]. If $\gamma_3 \subset G$ is a Jordan half-interval determining an accessible point of Γ distinct from the points determined by γ_1 and γ_2, then after prior deletion of an initial part of γ_3 (if necessary), we can assume that γ_3 and $\tilde{\gamma}$ are disjoint, so that either $\gamma_3 \subset g_1$ or $\gamma_3 \subset g_2$. Therefore γ_3 determines an accessible boundary point of one of the domains g_1 and g_2. Let H be the set of all accessible points of Γ. Then, given any two distinct points

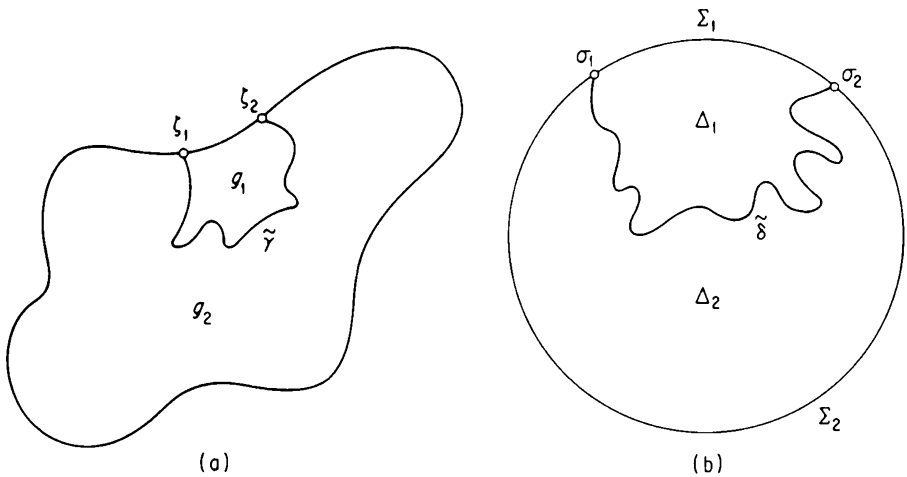


FIGURE 2.16

$\eta_1, \eta_2 \in H$,[15] every other point of H can be regarded as an accessible point either of the boundary of g_1 or of the boundary of g_2 (but not of both). Let H_1 be the set consisting of η_1, η_2 and all points of H assigned in this way to the boundary of g_1, and let H_2 be the set consisting of η_1, η_2 and all points of H assigned to the boundary of g_2. Then the sets H_1 and H_2, called *H-intervals* (or simply *intervals*, when there is no possibility of confusion),[16] are said to be *complementary* in the sense that $H = H_1 \cup H_2$, and the points η_1 and η_2 are called the *end points* of H_1 and H_2. Under a conformal mapping $w = f(z)$ carrying G into K, every point of H_1 goes into a point of Σ_1, and every point of H_2 goes into a point of Σ_2. Since the decomposition $\Sigma = \Sigma_1 \cup \Sigma_2$ is uniquely determined by the end points σ_1 and σ_2 of the arcs

[15] Note the need to denote points of H by symbols different from those used to denote final points of Jordan half-intervals (representing points of H).

[16] Less tersely, H-intervals are intervals of the set of accessible points of the boundary Γ. Note that H-intervals are the analogues of *closed* intervals.

Σ_1 and Σ_2, the decomposition $H = H_1 \cup H_2$ is uniquely determined by the end points η_1 and η_2 of the intervals H_1 and H_2, and does not depend on the choice of the disjoint half-intervals γ_1 and γ_2 determining η_1 and η_2 (or on the Jordan arc γ used to connect the initial points of γ_1 and γ_2), a fact which justifies thinking of H_1 and H_2 as "intervals." The two domains g_1 and g_2 whose sets of accessible boundary points contain H_1 and H_2, respectively, are said to be *adjacent* to H_1 and H_2. In this way, under the conformal mapping $w = f(z)$, a pair of complementary circular arcs Σ_1 and Σ_2 is associated with every pair of complementary H-intervals H_1 and H_2, and a pair of domains Δ_1 and Δ_2 adjacent to Σ_1 and Σ_2 is associated with every pair of domains g_1 and g_2 adjacent to H_1 and H_2. In fact, it is easy to see that there is a one-to-one correspondence between H-intervals and circular arcs whose end points are images of accessible points.

A sequence of H-intervals $\{H^{(n)}\}$ is said to be *nested* if for every n, $H^{(n+1)}$ is contained in $H^{(n)}$ and both end points of $H^{(n+1)}$ are different from those of $H^{(n)}$. We define nested sequences of arcs of a circle (or, more generally, of any Jordan curve) in exactly the same way. A nested sequence of H-intervals $\{H^{(n)}\}$ is said to be *shrinking* if $\bigcap H^{(n)}$, the intersection of all the intervals $H^{(n)}$ $(n = 1, 2, \ldots)$, does not contain two distinct accessible points.

THEOREM 2.13. *Given a shrinking nested sequence of H-intervals* $\{H^{(n)}\}$, *let* $\{\Sigma^{(n)}\}$ *be the sequence of arcs of the unit circle* Σ *corresponding to* $\{H^{(n)}\}$ *under a function* $w = f(z)$ *mapping* G *conformally onto the unit disk. Then* $\{\Sigma^{(n)}\}$ *is a nested sequence, and* $\bigcap\Sigma^{(n)}$, *the intersection of all the arcs* $\Sigma^{(n)}$ $(n = 1, 2, \ldots)$ *contains a unique point* $w_0 \in \Sigma$.

Proof. It is obvious that the sequence $\{\Sigma^{(n)}\}$ is nested. Every $\Sigma^{(n)}$ is a closed arc and hence a continuum.[17] Therefore, according to Theorem I.4.17, $\bigcap\Sigma^{(n)}$ is a continuum contained in Σ. But $\bigcap\Sigma^{(n)}$ cannot be an arc, since then, according to Theorem 2.11, $\bigcap H^{(n)}$ would contain infinitely many accessible points, contrary to hypothesis. Therefore $\bigcap\Sigma^{(n)}$ must be a set containing a single point $w_0 \in \Sigma$.

COROLLARY. *If* w_0 *is the image of an accessible point* η, *then* $\bigcap H^{(n)}$ *is the set containing only* η. *Otherwise,* $\bigcap H^{(n)}$ *is the empty set.*

DEFINITION 1. *Any shrinking nested sequence of H-intervals* $\{H^{(n)}\}$ *is said to determine a prime end (of the domain G), and the point* w_0 *figuring in Theorem* 2.13 *is called the image of the prime end under the mapping* $w = f(z)$. *Prime ends are said to be distinct if they have distinct images and the same if they have the same image.*

---

[17] Every continuum contained in Σ is either a closed arc or a set containing only one point (see Problem I.4.9).

THEOREM 2.14. *Let $\{H^{(n)}\}$ and $\{\tilde{H}^{(n)}\}$ be two shrinking nested sequences of H-intervals. Then $\{H^{(n)}\}$ and $\{\tilde{H}^{(n)}\}$ determine the same prime end if and only if every interval $H^{(m)}$ contains all the intervals $\tilde{H}^{(n)}$ starting from some value of n.*

Proof. Let $\{\Sigma^{(n)}\}$ and $\{\tilde{\Sigma}^{(n)}\}$ be the nested sequences of arcs of the unit circle Σ corresponding to $\{H^{(n)}\}$ and $\{\tilde{H}^{(n)}\}$, respectively. Then

$$\bigcap \Sigma^{(n)} = \{w_0\}, \qquad \bigcap \tilde{\Sigma}^{(n)} = \{\tilde{w}_0\},$$

where w_0 and $\tilde{w}_0$ are two points of Σ. But

$$H^{(m)} \supset \tilde{H}^{(n)} \qquad \text{for all } n > N(m) \text{ and } m = 1, 2, \ldots \tag{2.22}$$

implies that

$$\Sigma^{(m)} \supset \tilde{\Sigma}^{(n)} \qquad \text{for all } n > N(m) \text{ and } m = 1, 2, \ldots \tag{2.23}$$

Therefore

$$\bigcap \Sigma^{(n)} = \bigcap \tilde{\Sigma}^{(n)}, \tag{2.24}$$

and hence $w_0 = \tilde{w}_0$. Conversely, $w_0 = \tilde{w}_0$ implies (2.24) and hence (2.23) and (2.22).

COROLLARY. *The roles of the sequences $\{H^{(n)}\}$ and $\{\tilde{H}^{(n)}\}$ can be interchanged in the condition figuring in Theorem* 2.14.

Proof. The condition (2.24) holds if and only if $\tilde{\Sigma}^{(n)} \supset \Sigma^{(m)}$ and hence $\tilde{H}^{(n)} \supset H^{(m)}$ for all $m > M(n)$ and $n = 1, 2, \ldots$.

THEOREM 2.15. *Given any point $w_0 \in \Sigma$, there is a prime end of G with image w_0.*

Proof. Given any $w_0 \in \Sigma$, it follows from Theorem 2.11 that there is a nested sequence $\{\Sigma^{(n)}\}$ (in fact, infinitely many) of arcs of Σ, whose end points are images of accessible points of the boundary of G, such that

$$\bigcap \Sigma^{(n)} = \{w_0\}$$

(whether or not w_0 itself is the image of an accessible point). For every n, let $H^{(n)}$ be the H-interval corresponding to $\Sigma^{(n)}$ under the mapping $z = f^{-1}(w)$. Then $\{H^{(n)}\}$ is a shrinking nested sequence determining a prime end τ_1 with image w_0 in the sense of Definition 1.

COROLLARY. *Every conformal mapping $w = f(z)$ of a bounded simply connected domain G onto the unit disk $|w| < 1$ establishes a one-to-one correspondence between the prime ends of G and the points of the unit circle $|w| = 1$.*

DEFINITION 2. *Let τ_1 be a prime end of G, determined by a sequence $\{H^{(n)}\}$, and given any n, let $g^{(n)} \subset G$ be a domain adjacent to $H^{(n)}$. Then*

the sequence $\{g^{(n)}\}$ *is said to be a chain of domains tending to* η *if* $g^{(n+1)} \subset g^{(n)}$ *for every* n *and if* $\bigcap \bar{g}^{(n)}$, *the intersection of all the closed domains* $\bar{g}^{(n)}$ $(n = 1, 2, \ldots)$, *contains no points of* G.

The introduction of Definition 2 is justified by

THEOREM 2.16. *If* η *is a prime end of* G, *then there exists a chain of domains tending to* η.

Proof. Let $w = f(z)$ be a conformal mapping of G onto the unit disk, and let $\{\Sigma^{(n)}\}$ be the sequence of circular arcs corresponding to some sequence $\{H^{(n)}\}$ determining η. Let $\gamma_1^{(n)}$ and $\gamma_2^{(n)}$ be two Jordan half-intervals determining the end points of the interval $H^{(n)}$, and let $\delta_1^{(n)}$ and $\delta_2^{(n)}$ be their images under $w = f(z)$. Construct the domain $\Delta^{(n)}$ bounded by the arc $\Sigma^{(n)}$ of the unit circle, by an arc of the concentric circle $|w| = r_n < 1$, and by the two parts of the half-intervals $\delta_1^{(n)}$ and $\delta_2^{(n)}$ remaining after deletion of certain initial parts (see Figure 2.17). We impose the further condition that

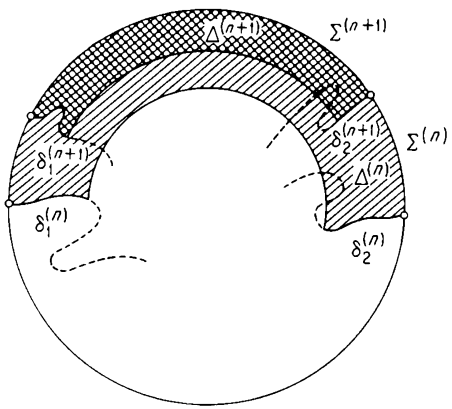


FIGURE 2.17

$$r_n > 1 - \frac{1}{n}.$$

Since $\Sigma^{(n+1)} \subset \Sigma^{(n)}$ and both end points of $\Sigma^{(n+1)}$ are different from those of $\Sigma^{(n)}$, the half-intervals $\delta_1^{(n+1)}$ and $\delta_2^{(n+1)}$ are contained in $\Delta^{(n)}$, at least after deletion of certain initial parts. Choosing $r_{n+1} > r_n$ and

$$r_{n+1} > 1 - \frac{1}{n+1}$$

such that the circle $|w| = r_{n+1}$ intersects both $\delta_1^{(n+1)}$ and $\delta_2^{(n+1)}$, we construct a domain $\Delta^{(n+1)}$ in exactly the same way as Δ_n, this time using $\Sigma^{(n+1)}$, an arc of the circle $|w| = r_{n+1}$ and parts of the half-intervals $\delta_1^{(n+1)}$ and $\delta_2^{(n+1)}$. Carrying out this construction for every $n = 1, 2, \ldots$, we obtain a sequence of domains $\Delta^{(n)}$ adjacent to the arcs $\Sigma^{(n)}$ such that $\Delta^{(n+1)}$ is contained both in $\Delta^{(n)}$ and in the annulus

$$1 - \frac{1}{n} < |w| < 1.$$

Thus, if

$$g^{(n)} = f^{-1}(\Delta^{(n)}) \qquad (n = 1, 2, \ldots),$$

then $g^{(n)}$ is adjacent to $H^{(n)}$ and $g^{(n+1)} \subset g^{(n)}$. Moreover, $\bigcap \bar{g}^{(n)}$ contains no points of G, since otherwise $\bigcap \bar{\Delta}^{(n)}$, the intersection of all the closed domains $\bar{\Delta}^{(n)}$ $(n = 1, 2, \ldots)$, would contain a point of the unit disk, contrary to the construction. Therefore $\{g^{(n)}\}$ is a chain of domains tending to η, as asserted.

THEOREM 2.17. *If $\{g^{(n)}\}$ is a chain of domains tending to a prime end η, then the set*

$$F = \bigcap \bar{g}^{(n)}$$

is a continuum containing only boundary points of G.

Proof. Since every $\bar{g}^{(n)}$ is a continuum and $\bar{g}^{(n+1)} \subset \bar{g}^{(n)}$ $(n = 1, 2, \ldots)$, the assertion follows from Theorem I.4.17 and the fact that, by definition, F contains no points of G.

DEFINITION 3. *Let η be a prime end of G, and let $\{g^{(n)}\}$ be any chain of domains tending to η. Then the points of the continuum $F = \bigcap \bar{g}^{(n)}$ are said to belong to η, and F is said to represent η.*[18]

In stating this definition, we rely on

THEOREM 2.18. *Let $\{g^{(n)}\}$ and $\{g'^{(n)}\}$ be two chains of domains tending to the same prime end η. Then*

$$\bigcap \bar{g}^{(n)} = \bigcap \bar{g}'^{(n)}, \tag{2.25}$$

i.e., the continuum $F = \bigcap \bar{g}^{(n)}$ is independent of the choice of the chain tending to η.

Proof. The chains $\{g^{(n)}\}$ and $\{g'^{(n)}\}$ correspond to two sequences $\{H^{(n)}\}$ and $\{H'^{(n)}\}$ of H-intervals determining η. Under conformal mapping of G onto the unit disk K, two sequences $\{\Sigma^{(n)}\}$ and $\{\Sigma'^{(n)}\}$ of arcs of the unit circle Σ are associated with $\{H^{(n)}\}$ and $\{H'^{(n)}\}$, and two sequences $\{\Delta^{(n)}\}$ and $\{\Delta'^{(n)}\}$ of domains adjacent to these arcs are associated with $\{g^{(n)}\}$ and $\{g'^{(n)}\}$. Since $\{H^{(n)}\}$ and $\{H'^{(n)}\}$ both determine the same prime end η, we have

$$\bigcap \Sigma^{(n)} = \bigcap \Sigma'^{(n)} = \{w_0\}$$

for some $w_0 \in \Sigma$. Since $\bigcap \bar{\Delta}^{(n)}$ can contain no points of K, we have

$$\bigcap \bar{\Delta}^{(n)} = \bigcap \Sigma^{(n)} = \{w_0\},$$

and similarly,

$$\bigcap \bar{\Delta}'^{(n)} = \bigcap \Sigma'^{(n)} = \{w_0\},$$

and hence

$$\bigcap \bar{\Delta}^{(n)} = \bigcap \bar{\Delta}'^{(n)}.$$

---

[18] We do not identify F with η, since the same set F can represent more than one prime end.

It follows that $\Delta^{(m)}$ contains all the $\Delta'^{(n)}$ starting from some value of n. Therefore the same is true of $g^{(m)}$ and $g'^{(n)}$. But then

$$\bigcap \bar{g}^{(n)} \supset \bigcap \bar{g}'^{(n)}, \tag{2.26}$$

and similarly,

$$\bigcap \bar{g}'^{(n)} \supset \bigcap \bar{g}^{(n)}, \tag{2.27}$$

since the role of the primed and unprimed domains can be interchanged. Comparing (2.26) and (2.27), we obtain (2.25).

We are now in a position to analyze and classify the prime ends of an arbitrary bounded simply connected domain G. This is done by examining the properties of the continuum F representing a given prime end η. As will soon be apparent, there is an essential difference between the case where F contains only one point and the case where F contains more than one point (and hence infinitely many points).

THEOREM 2.19. *Let $w = f(z)$ be a conformal mapping of a bounded simply connected domain G with boundary Γ onto the unit disk K. Let η be a prime end of G, with image w_0 on the unit circle, and suppose η is represented by the continuum F. Then, if $\{w_n\}$ is any sequence of points of K converging to w_0, the sequence $\{z_n\} = \{f^{-1}(w_n)\}$ has all its limit points in F. Conversely, every point $z_0 \in F$ is the limit of a sequence $\{z_n\}$ of points of G such that $\{w_n\} = \{f(z_n)\}$ converges to w_0.*

Proof. Let $\{g^{(n)}\}$ be a chain of domains tending to η, and let $\{\Delta^{(n)}\}$ be the images of these domains under the mapping $w = f(z)$, so that

$$\bigcap \bar{g}^{(n)} = F, \qquad \bigcap \bar{\Delta}^{(n)} = \{w_0\}.$$

If $\{w_n\}$ converges to w_0, then, given any m, $w_n \in \Delta^{(m)}$ for all sufficiently large $n > N(m)$. Therefore $z_n = f^{-1}(w_n) \in g^{(m)}$ if $n > N(m)$, i.e., all the limit points of $\{z_n\}$ belong to $\bigcap \bar{g}^{(n)} = F$. On the other hand, if $z_0 \in F$, then z_0 is the limit of some sequence $\{z_n\}$, where $z_n \in g^{(n)}$, since the neighborhood $|z - z_0| < 1/n$ intersects all the $g^{(n)}$. Therefore $w_n = f(z_n) \in \Delta^{(n)}$, i.e., $\{w_n\}$ converges to w_0, as asserted.

COROLLARY 1. *If F contains only one point z_0, then*

$$\lim_{\substack{w \to w_0 \\ w \in K}} f^{-1}(w) = z_0. \tag{2.28}$$

On the other hand, if F contains more than one point, then

$$\lim_{n \to \infty} f^{-1}(w_n)$$

does not take the same value for every sequence $\{w_n\}$ of points of K converging to w_0, i.e., (2.28) does not exist.

COROLLARY 2. *If F contains only one point z_0, and if $\delta \subset K$ is an arbitrary Jordan half-interval with final point w_0, then $\gamma = f^{-1}(\delta) \subset G$ is a Jordan half-interval with final point z_0.*

Thus, if z_0 is the only point of F, the point w_0, which is defined as the image of the prime end η, is also the image of a unique accessible point (of the boundary Γ), represented by the point z_0. In this case, we identify η with the accessible point, and call η a *prime end of the first kind.* The point z_0 itself is said to *belong to* η or to *represent* η.

The situation is more complicated if F contains more than one point, since then the expression

$$\lim_{\substack{w \to w_0 \\ w \in K}} f^{-1}(w)$$

is meaningless. However, it may still happen that w_0 is the image of an accessible point of the boundary Γ, i.e., that there exists a Jordan half-interval $\delta \subset K$ with final point w_0 which is the image of a Jordan half-interval $\gamma \subset G$ with final point $z_0 \in F$. This means that

$$\lim_{\substack{w \to w_0 \\ w \in \delta}} f^{-1}(w) = z_0,$$

so that γ determines an accessible point represented by z_0. In this case, η is called a *prime end of the second kind,* and we say that the point z_0 *belongs to* η. Moreover, z_0 is called an *accessible point of* η, and every point of F other than z_0 is called an *inaccessible point of* η. Figure 2.18 shows a prime end η of the second kind, represented by the line segment AB. Here we also show two domains $g^{(n)}$ and $g^{(n+1)}$ of a chain tending to η. The point A is an accessible point of η, and all other points of AB are inaccessible points of η. It should be noted that the same points (except for A) represent prime ends of the first kind, as shown in the figure for the case of the point C.

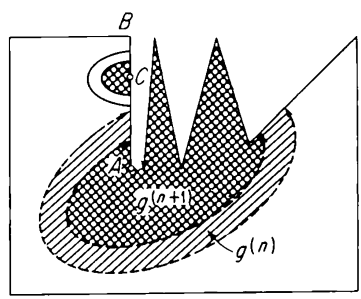


FIGURE 2.18

Finally, we consider the case where the prime end η contains no accessible points, i.e., where w_0 is not the image of an accessible point of Γ. Then, given any Jordan half-interval δ with final point w_0, the inverse image $\gamma = f^{-1}(\delta)$ will not be a Jordan half-interval, since

$$\lim_{\substack{w \to w_0 \\ w \in \delta}} f^{-1}(w)$$

does not exist. In any event, all the limit points of the curvilinear half-interval γ belong to the continuum F, and two subcases can be distinguished. If the limit set of every curvilinear half-interval $\gamma = f^{-1}(\delta)$ coincides with F, the element η is called a *prime end of the third kind.* On the other hand, it may be possible to find a δ such that the limit set of $\gamma = f^{-1}(\delta)$ is a proper subset of the continuum F, and then η is called a *prime end of the fourth kind.*

To recognize prime ends of either kind, there is no need to resort to the conformal mapping $z = f^{-1}(w)$. Instead, we need only choose a chain of domains $\{g^{(n)}\}$ tending to η, and then consider all possible curvilinear half-intervals γ, with limit sets in Γ, which are contained in every domain $g^{(n)}$ starting from some point of γ. (In fact, it is just these half-intervals which are the inverse images of all possible Jordan half-intervals δ with final points at w_0, the image of the prime end η.) The rest of the problem consists in comparing the limit sets of the half-intervals γ with the continuum F.

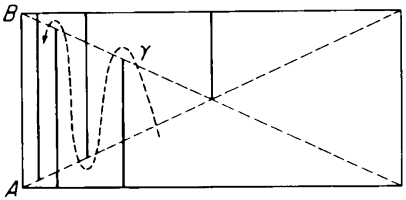


FIGURE 2.19

In Figure 2.19, the segment AB represents a prime end of the third kind, and the dashed curve shows a typical half-interval γ whose limit set necessarily coincides with AB. On the other hand, the segment AB in Figure 2.10 represents a prime end of the fourth kind, where this time the dashed curve shows a half-interval γ whose limit set is a proper subset of AB, namely the segment CD.

8. Boundary Behavior of Conformal Mappings

Let η be a prime end of a bounded simply connected domain G, and let $\{g^{(n)}\}$ be a chain of domains tending to η. Then a sequence $\{z_n\}$ of points of G is said to *converge to* η if z_n belongs to $g^{(m)}$ for all $n > N(m)$ and $m = 1, 2, \ldots$. Clearly, if this condition is satisfied for one chain of domains tending to η, it is also satisfied for any other chain of domains tending to η, a fact which justifies the definition.

THEOREM 2.20. *Suppose the function $w = f(z)$ maps G conformally onto the unit disk $K: |w| < 1$, thereby establishing a one-to-one correspondence between the prime ends of G and the points of the unit circle $\Sigma: |w| = 1$.*[19] *Then this correspondence is "continuous" in both directions in the sense that every sequence $\{z_n\}$ of points of G converging to a prime end η is mapped into a sequence $\{w_n\}$ of points of K converging to the point $w_0 \in \Sigma$ which is the image of η, and conversely.*

Proof. The sequence $\{w_n\}$ converges to w_0 if and only if $w_n \in \Delta^{(m)}$ for all $n > N(m)$ and $m = 1, 2, \ldots$, where $\{\Delta^{(n)}\}$ is any sequence of domains $\Delta^{(n)} \subset K$ such that $\bigcap \overline{\Delta}^{(n)} = \{w_0\}$. Therefore $\{w_n\}$ converges to w_0 if and only if $z_n = f^{-1}(w_n) \in g^{(m)} = f^{-1}(\Delta^{(m)})$ for all $n > N(m)$ and $m = 1, 2, \ldots$.

[19] See Theorem 2.15, Corollary.

Remark. In applying Theorem 2.20, it must be kept clearly in mind that convergence of a sequence of points $\{z_n\}$ to a prime end η does not imply convergence of $\{z_n\}$ in the usual sense. In fact, the set of limit points of $\{z_n\}$ may coincide with any nonempty closed subset E of the continuum F representing η. Only in the case where η is a prime end of the first kind, so that F contains only one point z_0, can we infer from the convergence of $\{z_n\}$ to η that $\{z_n\}$ converges to the point z_0. But even then, the converse fails to hold, i.e., a sequence $\{z_n\}$ converging to a point z_0 may not converge to a prime end represented by z_0. For example, consider a sequence $\{z_n\}$ whose points lie alternately on the half-intervals γ_1 and γ' shown in Figure 2.11, and suppose $\{z_n\}$ converges to the point ζ representing two different prime ends. Then $\{z_n\}$ cannot converge to either prime end. However, the subsequence of $\{z_n\}$ consisting only of points on γ_1 converges to one of these prime ends, and the subsequence consisting only of points on γ' converges to the other prime end.

THEOREM 2.21. *The function* $z = f^{-1}(w)$ *mapping the unit disk* $K: |w| < 1$ *conformally onto* G *is continuous on the closed disk* $\overline{K}: |w| \leqslant 1$ *(and hence uniformly continuous on* K *itself*[20]*) if and only if all the prime ends of the domain* G *are of the first kind.*[21]

Proof. Suppose every prime end of G is of the first kind, and let $\{w_n\}$ be a sequence of points of K converging to a point w_0 of the unit circle $\Sigma: |w| = 1$. Then, according to Theorem 2.19, Corollary 1, the sequence $\{z_n\} = \{f^{-1}(w_n)\}$ converges to a point z_0 of the boundary of G. We define the function $f^{-1}(w)$ on the closed disk $\overline{K}$ by setting $f^{-1}(w_0) = z_0$ (and similarly for every point of Σ). Then $f^{-1}(w)$ is continuous on $\overline{K}$. In fact, suppose the opposite, so that there exists a sequence of points $\{w_n'\}$, $w_n' \in \overline{K}$ converging to $w_0 \in \Sigma$ whose limit

$$\lim_{n \to \infty} f^{-1}(w_n')$$

either fails to exist or is different from $f^{-1}(w_0)$. For every point $w_n' \in \Sigma$, we can find a point $w_n'' \in K$ such that $|f^{-1}(w_n') - f^{-1}(w_n'')| < 1/n$ and $|w_n' - w_n''| < 1/n$. Thus, setting $w_n'' = w_n'$ if $w_n' \in K$, we obtain a sequence $\{w_n''\}$ of points of K such that $\{w_n''\}$ converges to w_0, while

$$\lim_{n \to \infty} f^{-1}(w_n'')$$

either fails to exist or is different from $f^{-1}(w_0)$. But this contradicts the fact that

$$f^{-1}(w_0) = \lim_{\substack{w \to w_0 \\ w \in K}} f^{-1}(w) = z_0,$$

thereby proving that $f^{-1}(w)$ is continuous on $\overline{K}$, as asserted.

---

[20] Cf. Problem I.3.26.

[21] Examples of such domains are given in Figures 2.11 and 2.12.

Conversely, suppose G has at least one prime end η which is not of the first kind, and let F be the continuum representing η. Then F contains more than one point, and hence we can find a sequence $\{w_n\}$ of points of K converging to a point $w_0 \in \Sigma$ such that the sequence $\{z_n\} = \{f^{-1}(w_n)\}$ has more than one limit point (these limit points belong to F), i.e., such that $\{z_n\}$ is divergent. Therefore, in this case, $f^{-1}(w)$ cannot be continuous on $\overline{K}$, and the proof is complete.

COROLLARY. *If all the prime ends of G are of the first kind, then the boundary of G is a closed continuous curve.*

Proof. If Γ is the boundary of G, then, according to Theorems I.6.3 and 2.21, Γ has the equation

$$z = f^{-1}(w) = f^{-1}(e^{it}) \qquad (0 \leqslant t \leqslant 2\pi),$$

where $f^{-1}(e^{it})$ is a continuous function of the parameter t.

DEFINITION. *A point of the boundary Γ belonging to only one prime end of the domain G is called a simple point of Γ.*

THEOREM 2.22. *The function $w = f(z)$ mapping G conformally onto the unit disk $K: |w| < 1$ is continuous on $\overline{G}$ (and hence uniformly continuous on G itself) if and only if all the boundary points of G are simple.*

Proof. Suppose $f(z)$ is continuous on $\overline{G}$, although some point z_0 in Γ (the boundary of G) belongs to at least two distinct prime ends η' and η''. (Why does every z_0 belong to *some* prime end?) Let $\{z_n'\}$ be a sequence of points of G converging to z_0 and simultaneously to η', and let $\{z_n''\}$ be a sequence converging to z_0 and simultaneously to η''. Then $\{w_n'\} = \{f(z_n')\}$ converges to a point w_0' of the unit circle Σ, while $\{w_n''\} = \{f(z_n'')\}$ converges to another point $w_0'' \in \Sigma$. But then

$$\lim_{\substack{z \to z_0 \\ z \in G}} f(z)$$

diverges, contrary to the hypothesis that $f(z)$ is continuous on $\overline{G}$. This contradiction shows that z_0 must be simple.

Conversely, let z_0 be a simple point of Γ, and let $\{z_n\}$ be an arbitrary sequence of points of G converging to z_0. Suppose $\{w_n\} = \{f(z_n)\}$ does not converge to a unique limit in Σ, so that $\{w_n\}$ has two (or more) distinct limit points w', $w'' \in \Sigma$, and let $\{w_n'\}$ and $\{w_n''\}$ be two subsequences of $\{w_n\}$ converging to w' and w'', respectively. Then the sequences $\{z_n'\} = \{f^{-1}(w_n')\}$ and $\{z_n''\} = \{f^{-1}(w_n'')\}$ both converge to the point z_0. However, at the same time, $\{z_n'\}$ converges to η' and $\{z_n''\}$ converges to η'', where η' and η'' are the distinct prime ends with images w' and w'',

respectively. Therefore z_0 belongs to both η' and η'', contrary to the assumption that z_0 is simple. This contradiction proves that

$$\lim_{\substack{z \to z_0 \\ z \in G}} f(z) = w_0.$$

Setting $f(z_0) = w_0$ and repeating this argument for every $z_0 \in \Gamma$, we obtain a function continuous on $\bar{G}$, as required.

COROLLARY. *The function $w = f(z)$ mapping G conformally onto the unit disk $K: |w| < 1$ establishes a homeomorphism between $\bar{G}$ and $\bar{K}$ if and only if all the prime ends of G are of the first kind and all the points of the boundary Γ are simple.*

Proof. Combine Theorems 2.21 and 2.22.

THEOREM 2.23. *Let G be a bounded simply connected domain with boundary Γ. Then all the prime ends of G are of the first kind and all the points of Γ are simple if and only if Γ is a closed Jordan curve.*

Proof. If all the prime ends of G are of the first kind and if all the points of Γ are simple, then it follows from the corollaries to Theorems 2.21 and 2.22 that Γ is a closed Jordan curve. Conversely, suppose Γ is a closed Jordan curve, let η be a prime end of G, and let $\{g^{(n)}\}$ be a chain of domains tending to η. Then the boundary of $g^{(n)}$ is a closed Jordan curve consisting of a Jordan arc $\Gamma^{(n)} \subset \Gamma$ (since Γ is a Jordan curve) and an open Jordan interval $\gamma^{(n)} \subset G$ with the same end points as $\Gamma^{(n)}$. Therefore the continuum F representing η equals $\bigcap \Gamma^{(n)}$, the intersection of all the nested arcs $\Gamma^{(n)}$ $(n = 1, 2, \ldots)$, i.e., F is either a set containing a single point of the boundary Γ or a Jordan arc $\Gamma' \subset \Gamma$. But the latter possibility is excluded, as we now show. In fact, if z_0 is a point of the arc Γ' other than its end points, then the distance $\rho(z_0, \Gamma'')$ between z_0 and the set $\Gamma'' = \Gamma - \Gamma'$ is a positive number (here we again use the fact that Γ is a Jordan curve). Let $z' \in G$ be such that $|z' - z_0| \leqslant \varepsilon < \rho$. Then obviously the nearest point $\zeta \in \Gamma$ belonging to the line segment joining z' and z_0 cannot belong to Γ'' and hence belongs to Γ'. Moreover, the rectilinear half-interval with initial point z' and final point ζ determines an accessible point of Γ, represented by the point $\zeta \in \Gamma'$, where the accessible point belongs to the boundary of every $g^{(n)}$ $(n = 1, 2, \ldots)$, since every $g^{(n)}$ contains the indicated half-interval except for some initial part. But we can find infinitely many such accessible points, since both $z_0 \in \Gamma'$ and the number ε $(0 \leqslant \varepsilon < \rho)$ are arbitrary. Therefore the prime end η contains infinitely many distinct accessible points, which is impossible. This contradiction shows that F contains only one point, i.e., η is a prime end of the first kind.

Finally, suppose η' and η'' are two different prime ends of G, and let $\{g'^{(n)}\}$ and $\{g''^{(n)}\}$ be two chains of domains tending to η' and η'', respectively. If

$$\Delta'^{(n)} = f(g'^{(n)}), \qquad \Delta''^{(n)} = f(g''^{(n)}),$$

then

$$\bigcap \Delta'^{(n)} = \{w'\}, \qquad \bigcap \Delta''^{(n)} = \{w''\},$$

where $w' \neq w''$. Therefore, for sufficiently large n, $\Delta'^{(n)} \cap \Delta''^{(n)} = 0$ and hence $g'^{(n)} \cap g''^{(n)} = 0$. But the boundary of $g'^{(n)}$ consists of two Jordan arcs with common end points, where one of these arcs $\Gamma'^{(n)}$ is contained in Γ, and the other $\gamma'^{(n)}$ is contained in G (except for its end points). Moreover, the point z_0' representing η' lies on $\Gamma'^{(n)}$ and is distinct from the end points of $\Gamma'^{(n)}$. Since $g'^{(n)} \cap g''^{(n)} = 0$, $g''^{(n)}$ must be contained in the subdomain of G bounded by the arc $\gamma'^{(n)}$ and the arc $\Gamma - \Gamma'^{(n)}$ (the latter arc has no points in common with $\Gamma'^{(n)}$ except its end points). Therefore the boundary of $g''^{(n)}$ consists of an arc $\Gamma''^{(n)} \subset \Gamma - \Gamma'^{(n)}$ and an arc $\gamma''^{(n)}$ contained in G (except for its end points). It follows that the point z_0' representing η' does not belong to the boundary of $g''^{(n)}$, and hence cannot represent η''. In other words, distinct prime ends are represented by distinct points, i.e., every point of Γ belongs to only one prime end. This completes the proof.

Thus a characteristic property of a closed Jordan curve Γ is the fact that every point of Γ represents a unique accessible boundary point (relative to the domain G bounded by Γ). We can summarize this property by saying that *all the points of a closed Jordan curve are simple and accessible from within.* Combining Theorems 2.22 and 2.23, we obtain the following basic

THEOREM 2.24 (*Boundary correspondence theorem*).[22] *Let G be a bounded simply connected domain with boundary Γ, and let $w = f(z)$ be any function mapping G conformally onto the disk $|w| < 1$. Then $w = f(z)$ establishes a homeomorphism between $\bar{G}$ and the closed disk $|w| \leqslant 1$, and hence between Γ and the circle $|w| = 1$, if and only if Γ is a closed Jordan curve.*[23]

COROLLARY. *Let G_1 and G_2 be the interiors of two closed Jordan curves Γ_1 and Γ_2, respectively. Then any conformal mapping of G_1 onto G_2 establishes a homeomorphism between Γ_1 and Γ_2.*

Proof. First map G_1 onto the disk $K: |w| < 1$, and then map K onto G_2.

Finally we examine the behavior of conformal mappings on suitably behaved subsets of Γ. Let G be an arbitrary simply connected domain,

[22] Already cited and used in Sec. II.27.

[23] In this case, G is called a *Jordan domain* (see p. 38).

with boundary Γ. Then a Jordan arc $\Gamma' \subset \Gamma$ is said to be an *accessible arc* of Γ if there exists a Jordan arc γ' with the same end points as Γ', contained in G except for its end points, such that the interior g of the Jordan curve $\gamma' \cup \Gamma'$ is contained in G. With the usual terminology, the domain g is said to be *adjacent* to Γ'. To clarify this concept, we observe that the line segment denoted by AB in Figure 2.18 is an accessible arc, but not the line segment denoted by AB in Figure 2.19.

THEOREM 2.25. *Let G be a simply connected (possibly unbounded) domain with boundary Γ, and let $w = f(z)$ be any function mapping G conformally onto the disk $|w| < 1$. Then $w = f(z)$ establishes a homeomorphism between any accessible Jordan arc $\Gamma' \subset \Gamma$ and an arc of the circle $|w| = 1$, provided that the values of $w = f(z)$ are considered only in a domain $g \subset G$ adjacent to Γ'.*

Proof. Under $w = f(z)$, the arc γ' figuring in the definition of the accessible arc Γ' goes into an open Jordan interval δ' with end points on the unit circle. (This follows from the fact that γ' determines two distinct accessible points of Γ.) The domain g bounded by the Jordan curve $\gamma' \cup \Gamma'$ is transformed into a domain Δ, bounded by the Jordan curve consisting of δ' and one of the arcs of the unit circle with the same end points as δ'. According to the corollary to Theorem 2.24, the function $w = f(z)$ mapping g conformally onto Δ establishes a homeomorphism between the boundaries of these two domains. Since the image of γ' is δ', the image of Γ' is an arc of the unit circle, and the proof is complete.

Example. To illustrate the results of this section, we consider the conformal mapping of the domain G shown in Figure 2.20 onto the unit disk. The boundary Γ of G consists of the two circles $L_1: |z| = 1$ and $L_2: |z| = 2$, together with a spiral $\left(\text{whose equation is } r = \frac{3}{2} + \frac{1}{\pi} \arctan \theta, \text{ say}\right)$ lying between L_1 and L_2, and approaching L_1 and L_2 "arbitrarily closely in both directions." Each point of the spiral represents two prime ends of the first kind, i.e., two accessible boundary points. On the other hand, each of the circles L_1 and L_2 represents a prime end of the third kind. To see this, consider the line segments shown in the figure, which lie along the positive real axis. Each of these segments is contained in G and divides the set of accessible boundary points of G into two complementary H-intervals, each with its own adjacent subdomain of G. One of these subdomains, called the

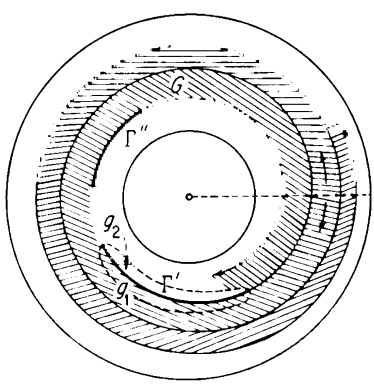


FIGURE 2.20

lower domain, winds away from the indicated segment in the clockwise direction and approaches L_1; the other subdomain, called the *upper domain*, winds in the counterclockwise direction and approaches L_2. Suppose we examine a sequence of segments approaching L_1, each time choosing the H-intervals associated with the lower domain. In this way, we obtain a shrinking nested sequence of H-intervals, determining a prime end of G, which is obviously represented by L_1, since L_1 is the intersection of all the closed lower domains. Similarly, we find that L_2 represents another prime end. Moreover, according to the definition on p. 65, both of these prime ends are of the third kind.

It follows that under conformal mapping of the domain G onto the unit disk, each point of the spiral is represented by two points of the unit circle $\Sigma: |w| = 1$, while each circle L_1 and L_2 is represented by one point of Σ. Since every arc Γ' of the spiral is an accessible Jordan arc with two disjoint adjacent domains $g_1 \subset G$ and $g_2 \subset G$, as shown in the figure, the conformal mapping carries Γ' into two disjoint arcs of Σ, in keeping with the fact that every point of Γ represents two distinct prime ends. We also note that two pairs of arcs of Σ are associated with every pair of arcs Γ' and Γ'' of the spiral, where the circular arcs have common points if and only if the same is true of Γ' and Γ''. From all this, it is clear that under conformal mapping, the whole spiral goes into a pair of complementary arcs of Σ with common end points A and B, where one end point is the image of the whole circle L_1 and the other is the image of the whole circle L_2. Since the boundary Γ contains points which are not simple (actually, only the points of L_1 and L_2 are simple), and since some of the prime ends of G are not of the first kind, the function $w = f(z)$ cannot be continuous on $\overline{G}$, and the function $z = f^{-1}(w)$ cannot be continuous on the closed disk $|w| \leqslant 1$. In fact, $w = f(z)$ is discontinuous at every point of the spiral, while $z = f^{-1}(w)$ has only two points of discontinuity, namely A and B.

Finally we prove a result which will be needed later in studying polynomial approximation on closed domains (see Sec. 12):

THEOREM 2.26. *Let $\{G_n\}$ be a decreasing sequence of bounded simply connected domains converging to its kernel $G_{z_0} = G$ which is the interior of a closed Jordan curve Γ. For every $n = 1, 2, \ldots$, let $w = f_n(z)$ be the function which maps G_n conformally onto the unit disk $K: |w| < 1$ and satisfies the conditions*

$$f_n(z_0) = 0, \qquad f_n'(z_0) > 0.$$

Moreover, let $w = f(z)$ be the function which maps G conformally onto K and satisfies the conditions

$$f(z_0) = 0, \qquad f'(z_0) > 0,$$

thereby establishing a homeomorphism between $\overline{G}$ and $\overline{K}$. Then $\{f_n(z)\}$ converges uniformly to $f(z)$ on $\overline{G}$.

Proof. According to Carathéodory's mapping theorem (Theorem 2.1), the sequence $\{f_n(z)\}$ converges uniformly to $f(z)$ inside G. Every function $f_n(z)$ is uniformly continuous on $\overline{G}$ (since $\overline{G} \subset G_n$), and moreover, as we now show, the sequence $\{f_n(z)\}$ is *equicontinuous* on $\overline{G}$ (cf. Problem I.17.29), i.e., given any $\varepsilon > 0$, there exists a number $\delta(\varepsilon) > 0$ such that

$$|f_n(z') - f_n(z'')| < \varepsilon$$

for all n if z' and z'' are any two points in $\overline{G}$ satisfying the inequality

$$|z' - z''| < \delta(\varepsilon).$$

Suppose this is not the case. Then there must exist a number $\varepsilon_0 > 0$, a sequence $\{n_k\}$ of positive integers, where $n_1 < \cdots < n_k < \cdots$, and two sequences of points $\{z'_k\}$, $\{z''_k\}$ in G such that

$$|f_{n_k}(z'_k) - f_{n_k}(z''_k)| \geqslant \varepsilon_0, \tag{2.29}$$

despite the fact that

$$\lim_{k \to \infty} (z'_k - z''_k) = 0.$$

(There is clearly no loss of generality in assuming $z'_k, z''_k \in G$ instead of $z'_k, z''_k \in \overline{G}$). Going over to subsequences if necessary, but keeping the same notation, we assume that $\{z'_k\}$ and $\{z''_k\}$ have limits (which are necessarily equal):

$$\lim_{k \to \infty} z'_k = \lim_{k \to \infty} z''_k = \zeta. \tag{2.30}$$

The point $\zeta \in \overline{G}$ must belong to Γ. In fact, if ζ were an interior point of $\overline{G}$, then the uniform convergence of the sequence $\{f_n(z)\}$ in a neighborhood of ζ would imply

$$\begin{aligned} |f_{n_k}(z'_k) - f_{n_k}(z''_k)| \leqslant\ & |f_{n_k}(z'_k) - f(z'_k)| \\ & + |f(z'_k) - f(z''_k)| + |f(z''_k) - f_{n_k}(z''_k)| < \varepsilon_0 \end{aligned}$$

for all sufficiently large k, contrary to (2.29).

Again going over to subsequences if necessary, we can assume that both sequences $\{w'_k\} = \{f_{n_k}(z'_k)\}$ and $\{w''_k\} = \{f_{n_k}(z''_k)\}$ converge to limits w' and w'', respectively, where obviously $|w' - w''| \geqslant \varepsilon_0$. Moreover, w' and w'' belong to the unit circle $|w| = 1$. In fact, suppose that $|w'| < 1$, say. Then the function $z = f^{-1}(w)$ maps some neighborhood of the point w' into an arbitrarily small neighborhood of the point $z' = f^{-1}(w') \in G$. But since the sequence $\{f_n^{-1}(w)\}$ is uniformly convergent in a neighborhood of w' (cf. Theorem 2.1), it follows that the points $z'_k = f_{n_k}^{-1}(w'_k)$ also belong to an arbitrarily small neighborhood of z', starting from some value of k, i.e.,

$$\lim_{k \to \infty} z'_k = z' \neq \zeta,$$

contrary to (2.30).

We now draw a circle $\sigma: |w| = r < 1$. If r is sufficiently small, $f^{-1}(\sigma)$ will be contained in an arbitrarily small neighborhood of the point z_0, and hence for some r and all sufficiently large n, the curves $f_n^{-1}(\sigma)$ will lie inside a circle $s \subset G$ with center z_0. Let $w'_k = f_{n_k}(z'_{n_k})$, $w''_k = f_{n_k}(z''_{n_k})$ and let δ'_k, δ''_k be the portions of the radii joining the point $w = 0$ to the points w'_k, w''_k which lie outside the circle σ (see Figure 2.21). Since for sufficiently large k, the points w'_k and w''_k lie outside σ in arbitrarily small neighborhoods of w' and w'', respectively, there must exist a number $\alpha > 0$ such that the distance between δ'_k and δ''_k exceeds α for all $k > k_0$. Under the mapping $z = f_{n_k}^{-1}(w)$, the segments δ'_k, δ''_k go into disjoint Jordan arcs γ'_k, γ''_k in the domain G_{n_k}, with initial points inside s and final points z'_k, z''_k.

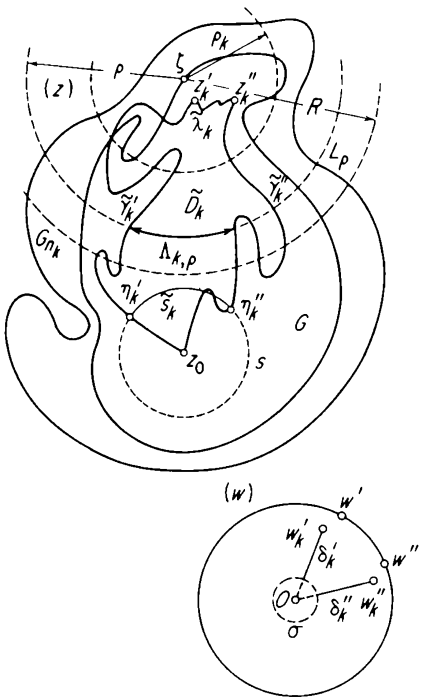


FIGURE 2.21

Next we observe that there exists a sequence of positive numbers $\{\rho_k\}$ converging to zero such that the points z'_k and z''_k belong to the disk $|z - \zeta| < \rho_k$ and can be joined by a Jordan arc λ_k contained in this disk and in the domain G. In fact, it follows from $z'_k, z''_k \in G$ and

$$\lim_{k\to\infty} z'_k = \lim_{k\to\infty} z''_k = \zeta$$

that

$$\lim_{k\to\infty} f(z'_k) = \lim_{k\to\infty} f(z''_k) = w_0.$$

But the points $f(z'_k)$ and $f(z''_k)$ can be joined by a straight line segment Δ_k contained in the unit disk $K: |w| < 1$, and hence the inverse image $\lambda_k = f^{-1}(\Delta_k)$ joins z'_k and z''_k, and is contained in G. Moreover, since $f^{-1}(w)$ is continuous on $\overline{K}$ (see Theorems 2.21 and 2.23, recalling that Γ is a closed Jordan curve), then, for sufficiently large k, the arc λ_k will be contained in an arbitrarily small neighborhood of ζ. This proves the existence of the required sequence $\{\rho_k\}$.

Now let the arc $\tilde{\lambda}_k$ be the part of λ_k going from the point ζ'_k in which λ_k last intersects γ'_k to the point ζ''_k in which λ_k first intersects γ''_k (in Figure 2.21, $\zeta'_k = z'_k$ and $\zeta''_k = z''_k$). Moreover, let the arc $\tilde{\gamma}'_k$ be the part

of γ'_k going from the point η'_k in which γ'_k last intersects the circle s to the point ζ'_k, and let the arc $\tilde{\gamma}''_k$ and the point η''_k be defined similarly. Finally, let the arc $\tilde{s}_k$ be the part of the circle s going from η'_k to η''_k. Then $-\tilde{\gamma}'_k + \tilde{s}_k + \tilde{\gamma}''_k - \tilde{\lambda}_k$ is a closed Jordan curve contained in G_{n_k}, whose interior $\tilde{D}_k$ is also contained in G_{n_k} (recall that G_{n_k} is simply connected). Every circle $L_\rho: |z - \zeta| = \rho$, where $\rho_k < \rho < R$ and R is the distance between ζ and $\tilde{s}_k$, intersects $\tilde{D}_k$ in some set of arcs, at least one of which, say $\Lambda_{k,\rho}$, joins a point $z'_{k,\rho} \in \tilde{\gamma}'_k$ to a point $z''_{k,\rho} \in \tilde{\gamma}''_k$. To verify the existence of such an arc $\Lambda_{k,\rho}$ (for every $k = 1, 2, \ldots$), we join any point of $\tilde{s}_k$ to a point of $\tilde{\lambda}_k$ by a Jordan arc $\tilde{\tau}_k$ contained in $\tilde{D}_k$ except for its end points. This is possible since the boundary of $\tilde{D}_k$ is a Jordan curve, and all points of a Jordan curve are accessible from within (see p. 70). Since the initial point of $\tilde{\tau}_k$ lies outside the circle L_ρ and the final point of $\tilde{\tau}_k$ lies inside L_ρ, the arc $\tilde{\tau}_k$ must intersect L_ρ in at least one of its arcs contained in $\tilde{D}_k$ (except for their end points). Among these arcs, there may be some with both end points on the same arc $\tilde{\gamma}'_k$ (or $\tilde{\gamma}''_k$); such arcs will be called *coterminal*. Every coterminal arc, together with an arc of $\tilde{\gamma}'_k$ (or $\tilde{\gamma}''_k$) with the same end points, forms a closed Jordan curve bounding some subdomain of $\tilde{D}_k$. The closure of this subdomain cannot contain the final point ζ of the arc $\tilde{\tau}_k$, and hence $\tilde{\tau}_k$ must eventually leave the subdomain on its way to the final point ζ. Clearly, the points of $\tilde{\tau}_k$ immediately preceding the first point of intersection of $\tilde{\tau}_k$ with a coterminal arc lie on the same side of L_ρ as the points of $\tilde{\tau}_k$ immediately following its last point of intersection with the arc, i.e., both sets of points belong to either the interior or the exterior of L_ρ. Therefore, if all the arcs of L_ρ contained in $\tilde{D}_k$ were coterminal, $\tilde{\tau}_k$ would never arrive at its final point ζ, since then $\tilde{\tau}_k$ would certainly approach its first point of intersection with a coterminal arc from outside L_ρ and hence would have to remain outside L_ρ after its last point of intersection with a coterminal arc. It follows that, as asserted, L_ρ contains at least one arc $\Lambda_{k,\rho}$ which is not coterminal and is intersected by $\tilde{\tau}_k$.

Next let the arc τ_k be the part of $\tilde{\tau}_k$ going from the point in which $\tilde{\tau}_k$ last intersects the circle $|z - \zeta| = R$ to the point in which $\tilde{\tau}_k$ first intersects the circle $|z - \zeta| = \rho_k$. Among all the subdomains of $\tilde{D}_k$ making up the intersection of $\tilde{D}_k$ with the annulus $\rho_k < |z - \zeta| < R$, let D_k be the one containing the arc τ_k, except for its end points. Obviously, τ_k intersects every $\Lambda_{k,\rho}$, where $\rho_k < \rho < R$ (since τ_k does), and hence every $\Lambda_{k,\rho}$ is contained in D_k (except for its end points). Since

$$f_{n_k}(z'_{k,\rho}) = w'_{k,\rho} \in \delta'_k, \qquad f_{n_k}(z''_{k,\rho}) = w''_{k,\rho} \in \delta''_k,$$

we find that

$$0 < \alpha < |w'_{k,\rho} - w''_{k,\rho}| = \left| \int_{z'_{k,\rho}}^{z''_{k,\rho}} f'_{n_k}(z)\, dz \right| \leqslant \int_{\Lambda_{k,\rho}} |f'_{n_k}(\zeta + \rho e^{i\theta})| \rho\, d\theta$$

(where f' is the derivative of f), and therefore

$$\alpha^2 < \int_{\Lambda_{k,\rho}} |f'_{n_k}(\zeta + \rho e^{i\theta})|^2 \rho^2 \, d\theta \int_{\Lambda_{k,\rho}} d\theta < 2\pi \int_{\Lambda_{k,\rho}} |f'_{n_k}(\zeta + \rho e^{i\theta})|^2 \rho^2 \, d\theta, \tag{2.31}$$

because of Schwarz's inequality

$$\left(\int_a^b f(x)g(x) \, dx \right)^2 \leqslant \int_a^b f^2(x) \, dx \int_a^b g^2(x) \, dx.$$

Multiplying both sides of (2.31) by $d\rho/\rho$ and integrating from ρ_k to R, we obtain

$$\begin{aligned} \alpha^2 \ln \frac{R}{\rho_k} &< 2\pi \int_{\rho_k}^{R} \rho \, d\rho \int_{\Lambda_{k,\rho}} |f'_{n_k}(\zeta + \rho e^{i\theta})|^2 \, d\theta \\ &\leqslant 2\pi \iint_{D_k} |f'_{n_k}(\zeta + \rho e^{i\theta})|^2 \rho \, d\rho \, d\theta. \end{aligned}$$

But

$$\iint_{D_k} |f'_{n_k}(\zeta + \rho e^{i\theta})|^2 \rho \, d\rho \, d\theta$$

is the area of the image of the domain D_k under the mapping $w = f_{n_k}(z)$, and hence does not exceed π. Thus we have finally arrived at the inequality

$$\alpha^2 \ln \frac{R}{\rho_k} < 2\pi^2,$$

which is impossible, since $\rho_k \to 0$ as $k \to \infty$. This contradiction with (2.29) proves that the sequence $\{f_n(z)\}$ is equicontinuous on $\overline{G}$, as asserted.

The rest of the proof is now straightforward. Given any $\varepsilon > 0$, we choose a number $\delta(\varepsilon) > 0$ such that

$$|f_n(z') - f_n(z'')| < \frac{\varepsilon}{3} \qquad (n = 1, 2, \ldots)$$

and

$$|f(z') - f(z'')| < \frac{\varepsilon}{3}$$

for any pair of points $z', z'' \in \overline{G}$ satisfying the condition $|z' - z''| < \delta(\varepsilon)$. Then let $\bar{g} \subset \overline{G}$ be a closed domain such that the distance between every point $z \in \overline{G}$ and $\bar{g}$ is less than $\delta(\varepsilon)$. To construct $\bar{g}$, we first find a number $\mu(\varepsilon)$ such that

$$|f^{-1}(w') - f^{-1}(w'')| < \delta(\varepsilon)$$

for every pair of points $w', w'' \in \overline{K}$ satisfying the condition $|w' - w''| < \mu(\varepsilon)$, and then choose g to be the inverse image under $w = f(z)$ of the disk $|w| < 1 - \mu(\varepsilon)$. In fact, if $a \in \overline{G} - \bar{g}$ and $b = f(a)$, then the point

$$\beta = \frac{b}{|b|} [1 - \mu(\varepsilon)]$$

lies on the circle $|w| = 1 - \mu(\varepsilon)$, and moreover $|b - \beta| < \mu(\varepsilon)$. Therefore, noting that $\alpha = f^{-1}(\beta) \in \bar{g}$ (why?), we obtain

$$|a - \alpha| = |f^{-1}(b) - f^{-1}(\beta)| < \delta(\varepsilon),$$

i.e., the distance between the point a and the set $\bar{g}$ is less than $\delta(\varepsilon)$. Because of the uniform convergence of $\{f_n(z)\}$ on the compact set $\bar{g}$, we have

$$|f(\zeta) - f_n(\zeta)| < \frac{\varepsilon}{3}$$

for all $\zeta \in \bar{g}$ and $n > N(\varepsilon)$. Moreover, as noted above, if z is any point of $\bar{G}$ and if ζ is any point of $\bar{g}$ whose distance from z is less than $\delta(\varepsilon)$, then

$$|f_n(z) - f_n(\zeta)| < \frac{\varepsilon}{3}, \qquad |f(z) - f(\zeta)| < \frac{\varepsilon}{3}.$$

Therefore

$$|f(z) - f_n(z)| \leqslant |f(z) - f(\zeta)| + |f(\zeta) - f_n(\zeta)| + |f_n(\zeta) - f_n(z)| < \varepsilon$$

if $n > N(\varepsilon)$. In other words, the sequence $\{f_n(z)\}$ converges uniformly to $f(z)$ on $\bar{G}$, and the proof is complete.

PROBLEMS

2.1. Prove that a kernel of a sequence of simply connected domains is itself simply connected.

Hint. Use Theorem I.4.14.

2.2. Let G_n and $G'_{n,\infty}$ be the same as in the proof of Theorem 2.2. Show that $G_n = (\bar{G}'_{n,\infty})^c$.

2.3. Let F be a continuum, with complement F^c. Prove that every component of F^c is simply connected.

2.4. What can go wrong with the proof of Theorem 2.2 if ρ is defined as the distance between $\bar{G}$ and z' rather than the distance between $\bar{G}$ and an arc γ joining z' to infinity?

2.5. Prove the converse of Theorem 2.3, i.e., if G is a kernel of a decreasing sequence of bounded simply connected domains, then G is a Carathéodory domain.

2.6. Prove that if G is a kernel of a decreasing sequence of bounded simply connected domains $\{G_n\}$, then $\{G_n\}$ converges to G (relative to any point in G).

2.7. Let G be a simply connected domain whose boundary contains more than one point, and let $f(z)$ be analytic and bounded on G, with zeros at the points

$\zeta_1, \ldots, \zeta_n, \ldots$. Let $w = \varphi(z)$ be the function mapping G conformally onto the unit disk $|w| < 1$, and suppose the series

$$\sum_{n=1}^{\infty} (1 - |\varphi(\zeta_n)|)$$

diverges. Prove that $f(z) \equiv 0$.

Hint. Recall Theorem II.8.6, Corollary and Theorem II.8.7.

2.8. Show that the second condition in the definition on p. 52 is equivalent to the following condition: If the neighborhood $\mathcal{N}(\zeta)$ is sufficiently small, then $\mathcal{N}(\zeta) \cap \gamma_1$ and $\mathcal{N}(\zeta) \cap \gamma_2$ are both contained in the same connected component of $\mathcal{N}(\zeta) \cap G$.

2.9. Prove that two distinct accessible boundary points η_1, η_2 of a domain G with boundary Γ have the following property: Let $\gamma_1, \gamma_2 \subset G$ be any two Jordan half-intervals determining η_1, η_2 and let the distance between each of two points $z_1 \in \gamma_1$ and $z_2 \in \gamma_2$ and the boundary Γ be less than $h > 0$. Then, for sufficiently small h, the set-theoretic diameter (see Problem I.3.27) of every continuous curve $L \subset G$ joining z_1 and z_2 exceeds some number $d > 0$. What is the geometric interpretation of this property in Figures 2.11 and 2.12?

2.10. Use the result of the preceding problem to define a distance between accessible boundary points η_1, η_2 which is nonzero for distinct η_1, η_2, even if η_1 and η_2 are represented by the same point of Γ.

2.11. Prove that any simply connected domain whose boundary contains more than one point can be mapped conformally onto a rectangle in such a way that four given accessible boundary points go into the vertices of the rectangle.

Hint. Recall Sec. II.19.

2.12. Give an example of a bounded simply connected domain G with a boundary point belonging to infinitely many distinct prime ends.

2.13. Given a closed Jordan curve Γ and any $\varepsilon > 0$, prove that there are closed Jordan curves Γ_1 and Γ_2 such that

1. $I(\Gamma_1) \supset \Gamma$ and $I(\Gamma) \supset \Gamma_2$;
2. $\mathcal{N}(z; \varepsilon) \cap \Gamma_1 \neq 0$ for all $z \in \Gamma$ and $\mathcal{N}(z; \varepsilon) \cap \Gamma \neq 0$ for all $z \in \Gamma_1$;
3. $\mathcal{N}(z; \varepsilon) \cap \Gamma_2 \neq 0$ for all $z \in \Gamma$ and $\mathcal{N}(z; \varepsilon) \cap \Gamma \neq 0$ for all $z \in \Gamma_2$.

Hint. Use Theorem 2.24.

Comment. In other words, Γ can be arbitrarily well approximated (cf. p. I.379) both by an "outside" Jordan curve Γ_1 and by an "inside" Jordan curve Γ_2.

2.14. Let Γ be a closed rectifiable Jordan curve with interior $I(\Gamma)$, and suppose $f(z)$ is analytic on $\overline{I(\Gamma)}$. Given any $z_0 \in I(\Gamma)$, prove that there are two distinct points $z_1, z_2 \in \Gamma$ such that

$$\frac{f(z_2) - f(z_1)}{z_2 - z_1} = f'(z_0).$$

Hint. According to Theorem II.4.5, the function $g(z) = f(z) - f'(z_0)z$ cannot be one-to-one on Γ.

2.15. Let Γ be a closed rectifiable Jordan curve, and let z_0 be any point in $I(\Gamma)$. Prove that z_0 is the midpoint of a line segment with end points on Γ.

Hint. Apply the result of the preceding problem to the function $f(z) = z^2$.

2.16. Let G be the interior of a closed *smooth* Jordan curve Γ, and let $w = f(z)$ be the function mapping G conformally onto the unit disk $K: |w| < 1$. Prove that $\arg f'(z)$ is continuous on $\bar{K}$.

Hint. Consider the function

$$g_\varepsilon(w) = \ln \frac{\varphi(e^{i\varepsilon}w) - \varphi(w)}{e^{i\varepsilon}w - w} \qquad (\varepsilon > 0),$$

where $\varphi(w) \equiv f^{-1}(w)$. Express $g_\varepsilon(w)$ in terms of the values of its imaginary part on the circle $|w| = 1$. Taking the limit as $\varepsilon \to 0$, express $\ln \varphi'(w)$ in terms of the inclination of the tangent to Γ at the point $z = \varphi(e^{i\theta})$. Then use the theory of Sec. II.24 to prove the continuity of $\arg \varphi'(w)$ and hence of $\arg f'(z)$.

Comment. To prove the continuity of $f'(z)$ itself, we need further conditions on Γ (cf. footnote 20, p. II.168).

3

APPROXIMATION BY RATIONAL FUNCTIONS AND POLYNOMIALS

9. Locally Analytic Functions

Let $f(z)$ be a single-valued function defined on a set E which is dense in itself.[1] Then $f(z)$ is said to be *locally analytic* on E if given any $z_0 \in E$, there is a neighborhood $\mathcal{N}(z_0)$ and a power series

$$\sum_{n=0}^{\infty} a_n(z - z_0)^n$$

such that

$$f(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n$$

for all $z \in \mathcal{N}(z_0) \cap E$. In the case where E is a domain, the concept of a locally analytic function on E reduces to that of a function which is single-valued and analytic on a domain.

Example. Suppose we divide the z-plane into infinitely many closed squares $\overline{\Delta}_1, \ldots, \overline{\Delta}_n, \ldots$ with sides of unit length parallel to the coordinate axes, and let $E = \Delta_1 \cup \cdots \cup \Delta_n \cup \cdots$ be the set of interior points of these squares. Then the function

$$f(z) = z^n \qquad (z \in \Delta_n)$$

is locally analytic on E.

[1] I.e., every point of E is a limit point of E (cf. p. I.324).

Now let $E = \mathcal{O}$ be an arbitrary open set. The set $\mathcal{O}$ is a domain if it is connected, but otherwise $\mathcal{O}$ is the union of countably many disjoint domains (see Theorem I.4.7). Given any integer $n > 0$, suppose we divide the z-plane into infinitely many closed squares with sides of length 3^{-n} parallel to the coordinate axes, where the origin is a vertex of one of the squares. Let $\overline{\Delta}$ be any of these squares such that $\overline{\Delta}$ and the eight squares bordering $\overline{\Delta}$ are all contained in $\mathcal{O}$. Then the union of all such squares minus its boundary is an

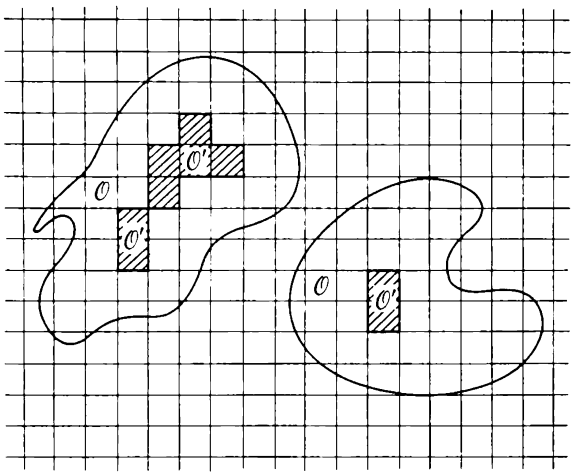


FIGURE 3.1

open set $\mathcal{O}_n'$ contained in $\mathcal{O}$ (see Figure 3.1). If $\mathcal{O}_n$ is the intersection of $\mathcal{O}_n'$ with the open square

$$-3^n < x < 3^n, \qquad -3^n < y < 3^n,$$

then $\mathcal{O}_n$ is a bounded open set such that $\overline{\mathcal{O}}_n \subset \mathcal{O}$. The boundary of $\mathcal{O}_n$ consists of a finite number of closed rectifiable Jordan curves, each of which is made up of a finite number of line segments. Moreover, it is clear that $\overline{\mathcal{O}}_n \subset \mathcal{O}_{n+1}$ and that given any compact set $F \subset \mathcal{O}$, we can find an integer $N(F) > 0$ such that $F \subset \mathcal{O}_n$ for all $n > N(F)$. These properties can be summarized by saying that $\{\mathcal{O}_n\}$ is an *increasing sequence of open sets approximating* $\mathcal{O}$.

THEOREM 3.1. *Let $\mathcal{O}$ be any open set, and let $f(z)$ be locally analytic on $\mathcal{O}$. Then there exists a sequence of rational functions $\{R_n(z)\}$ converging uniformly to $f(z)$ inside $\mathcal{O}$.*

Proof. Let $\{\mathcal{O}_n\}$ be the increasing sequence of open sets approximating $\mathcal{O}$ constructed above, and let Γ_n denote the boundary of $\mathcal{O}_n$ ($n = 1, 2, \ldots$). Then

$$f(z) = \frac{1}{2\pi i} \int_{\Gamma_{n+1}} \frac{f(\zeta)}{\zeta - z}\, d\zeta \tag{3.1}$$

for all $z \in \bar{\mathcal{O}}_n$, where the integral along Γ_{n+1} is understood to be a sum of integrals along the separate closed rectifiable Jordan curves forming Γ_{n+1}. This can be seen as follows: If z belongs to $\bar{\mathcal{O}}_n$, then z belongs to one of the (connected) components of $\mathcal{O}_{n+1}$. If this component $\mathcal{C}$ has boundary γ, then

$$f(z) = \frac{1}{2\pi i} \int_\gamma \frac{f(\zeta)}{\zeta - z}\, d\zeta. \tag{3.2}$$

On the other hand, the similar integrals along the boundaries of the remaining components of $\mathcal{O}_{n+1}$ vanish, giving rise to equations like (3.2) but with their left-hand sides equal to zero. Forming the sum of (3.2) and all these equations, we obtain (3.1), as required.

Now, given any $\varepsilon' > 0$, let $\delta(\varepsilon') > 0$ be such that $|f(\zeta') - f(\zeta'')| < \varepsilon'$ for any pair of points ζ', ζ'' belonging to Γ_{n+1} and satisfying the condition $|\zeta' - \zeta''| < \delta(\varepsilon')$. This is obviously possible, because of the continuity of $f(z)$ on Γ_{n+1}. Moreover, suppose we divide all the curves making up Γ_{n+1} into arcs σ_k, $k = 1, \ldots, N(\varepsilon')$, where σ_k has initial point $\zeta_k'^{(n+1)}$ and final point $\zeta_k''^{(n+1)}$, and the length λ_k of σ_k satisfies the condition

$$\lambda_k < \min\{\delta(\varepsilon'), \varepsilon'\}$$

for all $k = 1, \ldots, N(\varepsilon')$. The final point of the arc σ_k coincides with the initial point of the arc σ_{k+1}, except for the values of k for which σ_k and σ_{k+1} lie on different curves making up Γ_{n+1} (recall that in general Γ_{n+1} is a disconnected set).[2]

Next we approximate the integral (3.1) by the sum

$$S^{(n+1)}(z) = \frac{1}{2\pi i} \sum_{k=1}^{N(\varepsilon')} \frac{f(\zeta_k'^{(n+1)})}{\zeta_k'^{(n+1)} - z} (\zeta_k''^{(n+1)} - \zeta_k'^{(n+1)}), \tag{3.3}$$

corresponding to our partition of Γ_{n+1}. It is clear that

$$\begin{aligned}
&|f(z) - S^{(n+1)}(z)| \\
&\quad = \left| \frac{1}{2\pi i} \int_{\Gamma_{n+1}} \frac{f(\zeta)}{\zeta - z}\, d\zeta - \frac{1}{2\pi i} \sum_{k=1}^{N(\varepsilon')} \frac{f(\zeta_k'^{(n+1)})}{\zeta_k'^{(n+1)} - z} (\zeta_k''^{(n+1)} - \zeta_k'^{(n+1)}) \right| \\
&\quad = \left| \frac{1}{2\pi i} \sum_{k=1}^{N(\varepsilon')} \int_{\sigma_k} \left[\frac{f(\zeta)}{\zeta - z} - \frac{f(\zeta_k'^{(n+1)})}{\zeta_k'^{(n+1)} - z} \right] d\zeta \right| \\
&\quad = \left| \frac{1}{2\pi i} \sum_{k=1}^{N(\varepsilon')} \int_{\sigma_k} \left[\frac{f(\zeta) - f(\zeta_k'^{(n+1)})}{\zeta - z} + f(\zeta_k'^{(n+1)}) \frac{\zeta_k'^{(n+1)} - \zeta}{(\zeta - z)(\zeta_k'^{(n+1)} - z)} \right] d\zeta \right|.
\end{aligned}$$

Writing

$$M_{n+1} = \max_{z \in \Gamma_{n+1}} |f(z)|,$$

[2] Some of the closed curves making up Γ_{n+1} may share points (see Figure 3.1).

and denoting the distance between $\bar{\mathcal{O}}_n$ and Γ_{n+1} by $\rho_{n+1} > 0$, we find that

$$\begin{aligned}|f(z) - S^{(n+1)}(z)| &\leqslant \frac{1}{2\pi}\sum_{k=1}^{N(\varepsilon')}\int_{\sigma_k}\left|\frac{f(\zeta) - f(\zeta_k'^{(n+1)})}{\zeta - z}\right| d\sigma \\ &\quad + \frac{1}{2\pi}\sum_{k=1}^{N(\varepsilon')}\int_{\sigma_k}|f(\zeta_k'^{(n+1)})|\,\frac{|\zeta_k'^{(n+1)} - \zeta|}{|(\zeta - z)(\zeta'^{(n+1)} - z)|}\,d\sigma \\ &< \frac{1}{2\pi}\sum_{k=1}^{N(\varepsilon')}\varepsilon'\frac{\lambda_k}{\rho_{n+1}} + \frac{1}{2\pi}\sum_{k=1}^{N(\varepsilon')}M_{n+1}\frac{\varepsilon'\lambda_k}{\rho_{n+1}^2}. \end{aligned} \tag{3.4}$$

Since

$$\sum_{k=1}^{N(\varepsilon')}\lambda_k = \Lambda_{n+1},$$

where Λ_{n+1} is the sum of the lengths of all the curves making up Γ_{n+1}, (3.4) becomes

$$|f(z) - S^{(n+1)}(z)| < \frac{\varepsilon'\Lambda_{n+1}}{2\pi\rho_{n+1}^2}(\rho_{n+1} + M_{n+1}). \tag{3.5}$$

Obviously, the right-hand side of (3.5) can be made less than any preassigned $\varepsilon > 0$ by choosing ε' sufficiently small. Setting $\varepsilon = \varepsilon_n$, where $\varepsilon_n \to 0$ as $n \to \infty$, we obtain a sequence of rational functions $\{S^{(n+1)}(z)\}$, where $S^{(n+1)}(z)$ also depends on ε_{n+1}, which converges uniformly to $f(z)$ on every closed set $\bar{\mathcal{O}}_1, \bar{\mathcal{O}}_2, \ldots$ and hence uniformly to $f(z)$ inside $\mathcal{O}$. The proof is now completed by choosing $\{R_n(z)\} = \{S^{(n+1)}(z)\}$.

LEMMA. *Let F be a compact set with complement F^c, and let ζ be a point of F^c. Let $R(z)$ be a rational function whose only pole is at ζ, i.e.,*

$$R(z) = \frac{P(z)}{(z - \zeta)^k}, \qquad P(\zeta) \neq 0,$$

where the degree of the polynomial $P(z)$ does not exceed k. Suppose the point $\tilde{\zeta}$ belongs to the same component of F^c as ζ. Then, given any $\varepsilon > 0$, there exists a rational function

$$\tilde{R}(z) = \frac{\tilde{P}(z)}{(z - \tilde{\zeta})^{\tilde{k}}}, \qquad \tilde{P}(\tilde{\zeta}) \neq 0,$$

where the degree of the polynomial $\tilde{P}(z)$ does not exceed $\tilde{k}$, such that

$$|\tilde{R}(z) - R(z)| < \varepsilon \qquad (z \in F).$$

Proof. Let g be the component of F^c containing the points ζ and $\tilde{\zeta}$, let $\gamma \subset g$ be a Jordan curve joining ζ and $\tilde{\zeta}$, and let $\rho > 0$ be the distance between γ and F. Dividing γ into arcs $\sigma_1, \sigma_2, \ldots, \sigma_m$ of diameter less than

$\frac{1}{2}\rho$ (see Problem I.3.27), and denoting the points of division by $\zeta = \zeta_0$, $\zeta_1, \ldots, \zeta_m = \tilde{\zeta}$, we construct the rational function

$$R_1(z) = \frac{P(z)}{(z-\zeta)^k}\left[1 - \left(\frac{\zeta - \zeta_1}{z - \zeta_1}\right)^{n_1}\right]^k = \frac{P_1(z)}{(z-\zeta_1)^{n_1 k}},$$

where the degree of $P_1(z)$ does not exceed $k + n_1 k - k = n_1 k$. Then

$$\begin{aligned} |R_1(z) - R(z)| &= |R(z)|\left|\left[1 - \left(\frac{\zeta - \zeta_1}{z - \zeta_1}\right)^{n_1}\right]^k - 1\right| \\ &\leqslant M\left[\binom{k}{1}\left(\frac{\frac{1}{2}\rho}{\rho}\right)^{n_1} + \cdots\right] < M\frac{2^k}{2^{n_1}} \end{aligned}$$

for all $z \in F$, where

$$M = \max_{z \in F} |R(z)|.$$

Choosing n_1 sufficiently large, we find that

$$|R_1(z) - R(z)| < \frac{\varepsilon}{m} \qquad (z \in F).$$

By repetition of this argument, we obtain the required rational function $\tilde{R}(z) = R_m(z)$ at the mth step.

10. Functions Meromorphic on a Domain

We begin by using the above lemma to generalize some earlier results on partial fraction and infinite product expansions:

THEOREM 3.2 (*Generalization of Mittag-Leffler's theorem*).[3] *Let D be a domain in the extended plane which does not contain the point at infinity, and let*

$$\zeta_1, \ldots, \zeta_k, \ldots \tag{3.6}$$

be a sequence of distinct complex numbers all of whose limit points belong to the boundary of D. Moreover, let

$$G_1(z), \ldots, G_k(z), \ldots$$

be a sequence of rational functions of the form

$$G_k(z) = \frac{a^{(k)}_{-\beta_k}}{(z-\zeta_k)^{\beta_k}} + \cdots + \frac{a^{(k)}_{-1}}{z - \zeta_k} \qquad (k = 1, 2, \ldots), \tag{3.7}$$

where $\beta_k \neq 0$, $a^{(k)}_{-\beta_k} \neq 0$. Then there exists a function $f(z)$ which is analytic on D except for poles at the points (3.6), *and whose principal part at the pole ζ_k equals $G_k(z)$ for each $k = 1, 2, \ldots$*

---

[3] Theorem 3.2 reduces to Theorem II.10.10 (the earlier version of Mittag-Leffler's theorem) in the case where D is the finite plane, i.e., the domain whose boundary contains only the point ∞.

Proof. The restriction to domains which do not contain the point ∞ is not essential, but is made only for convenience.[4] Let $\{\mathcal{O}_n\}$ be the increasing sequence of open sets approximating D constructed on p. 81. Then it is easy to see that every component of $(\bar{\mathcal{O}})^c$, the complement of $\bar{\mathcal{O}}_n$, contains at least one boundary point of D (which may be the point ∞). This is obvious for the unbounded component, while if a bounded component contained no boundary points of D, it would be contained in D and hence in $\bar{\mathcal{O}}_n$ (why?), which is impossible.

Given any sequence $\{\varepsilon_n\}$ of positive numbers such that

$$\sum_{n=1}^{\infty} \varepsilon_n < \infty,$$

we now describe an inductive procedure for associating a certain rational function $R_n(z)$ with the set $\bar{\mathcal{O}}_n$ $(n = 1, 2, \ldots)$, with the aim of eventually showing that the desired function $f(z)$ is the sum of the series

$$R_1(z) + [R_2(z) - R_1(z)] + \cdots + [R_{n+1}(z) - R_n(z)] + \cdots. \quad (3.8)$$

To start the induction, consider all the points of the sequence $\{\zeta_k\}$ which belong to $\bar{\mathcal{O}}_1$. If there are no such points, we set $R_1(z) \equiv 0$. Otherwise, there are at most a finite number of such points, since if $\bar{\mathcal{O}}_1$ contained infinitely many points of $\{\zeta_k\}$, the set $\bar{\mathcal{O}}_1 \subset D$ would contain at least one limit point of $\{\zeta_k\}$, contrary to the hypothesis that all such limit points belong to the boundary of D. Then we define $R_1(z)$ as the sum of all the functions (3.7) with poles at these points.[5] Next consider all the points of $\{\zeta_k\}$ which belong to $\bar{\mathcal{O}}_2$ but not to $\bar{\mathcal{O}}_1$. If there are no such points, we set $R_2(z) \equiv R_1(z)$. Otherwise, there are at most a finite number m_1 of such points (by the same argument as before), which we denote by

$$\zeta_{k_1^{(1)}}, \ldots, \zeta_{k_{m_1}^{(1)}}.$$

Let the corresponding functions of the form (3.7) with poles at these points be denoted by

$$G_{k_1^{(1)}}(z), \ldots, G_{k_{m_1}^{(1)}}(z).$$

Using the lemma on p. 83, we approximate the function $G_{k_j^{(1)}}(z)$, $j = 1, \ldots, m_1$ by a rational function $r_j^{(1)}(z)$ such that 1) the only pole of $r_j^{(1)}(z)$ is at a boundary point of D belonging to the same component of $(\bar{\mathcal{O}}_1)^c$ as the point $\zeta_{k_j^{(1)}}$, and 2) the inequality

$$|G_{k_j^{(1)}}(z) - r_j^{(1)}(z)| < \frac{\varepsilon_1}{m_1}$$

---

[4] If $\zeta_k = \infty$ for some k, then the corresponding principal part is a polynomial rather than a rational function of the form (3.7).

[5] The construction of $R_1(z)$, unlike that of the functions $R_n(z)$, $n > 1$, does not involves the sequence $\{\varepsilon_n\}$.

holds for all $z \in \bar{\mathcal{O}}_1$. Then we define $R_2(z)$ by the formula

$$R_2(z) = R_1(z) + \sum_{j=1}^{m_1} [G_{k_j^{(1)}}(z) - r_j^{(1)}(z)].$$

Continuing this procedure, we construct $R_3(z)$ from $R_2(z)$ in the same way as $R_2(z)$ was constructed from $R_1(z)$, and so on. Thus

$$R_{n+1}(z) = R_n(z) + \sum_{j=1}^{m_n} [G_{k_j^{(n)}}(z) - r_j^{(n)}(z)] \qquad (n = 1, 2, \ldots),$$

where

$$|G_{k_j^{(n)}}(z) - r_j^{(n)}(z)| < \frac{\varepsilon_n}{m_n} \qquad (z \in \bar{\mathcal{O}}_n),$$

from which it is clear that all the poles of the function $R_{n+1}(z)$ in the domain G coincide with the points of the sequence $\{\zeta_k\}$ belonging to the set $\bar{\mathcal{O}}_{n+1}$, and moreover that the principal part of the Laurent expansion of $R_{n+1}(z)$ at each of these poles is given by the appropriate function (3.7). But the difference $R_{n+1}(z) - R_n(z)$, which has no poles in $\bar{\mathcal{O}}_n$, satisfies the inequality

$$|R_{n+1}(z) - R_n(z)| < \varepsilon_n$$

for all $z \in \bar{\mathcal{O}}_n$. It follows that the series (3.8) is uniformly convergent on $\bar{\mathcal{O}}_n$ if we exclude the first n terms which have poles in $\bar{\mathcal{O}}_n$. Choosing $f(z)$ to be the sum of the series (3.8), we see that

$$f(z) = R_n(z) + \sum_{k=n}^{\infty} [R_{k+1}(z) - R_k(z)] \qquad (z \in \bar{\mathcal{O}}_n)$$

for all $n = 1, 2, \ldots$, i.e., $f(z)$ can be represented on $\bar{\mathcal{O}}_n$ as the sum of an analytic function (here we use Theorem I.15.8) and a rational function whose poles in $\bar{\mathcal{O}}_n$ coincide with the points of $\{\zeta_k\}$ in $\bar{\mathcal{O}}_n$, with principal parts given by (3.7). Since $\bar{\mathcal{O}}_n$ eventually includes every point of D for sufficiently large n, the function $f(z)$ satisfies all the requirements of the theorem.

THEOREM 3.3 (*Generalization of Weierstrass' theorem*).[6] *Let D and $\{\zeta_k\}$ be the same as in Theorem* 3.2, *and let $\{\alpha_k\}$ be a sequence of positive integers. Then there exists a function $f(z)$ which is analytic on D and whose zeros coincide with the points of $\{\zeta_k\}$, where the zero at ζ_k is of order α_k for each $k = 1, 2, \ldots$*

[6] Theorem 3.3 reduces to Theorem II.10.1 (the earlier version of Weierstrass' theorem) in the case where D is the finite plane.

Proof. According to Theorem 3.2, there is a function $g(z)$ which is analytic on D except for poles at the points $\{\zeta_k\}$, and whose principal part at the pole ζ_k equals

$$\frac{\alpha_k}{z - \zeta_k} \qquad (k = 1, 2, \ldots).$$

Integration of $g(z)$ gives a multiple-valued function $\Phi(z)$ of the form

$$\Phi(z) = \alpha_k \operatorname{Ln}(z - \zeta_k) + \varphi_k(z) \qquad (k = 1, 2, \ldots)$$

in a neighborhood of ζ_k, where $\varphi_k(z)$ is the single-valued analytic function obtained by integrating the regular part of the Laurent expansion of $g(z)$ at ζ_k. It is easy to see that the function

$$f(z) = e^{\Phi(z)}$$

satisfies all the requirements of the theorem. In fact, $f(z)$ is obviously single-valued and analytic on D, and moreover the zeros of $f(z)$ coincide with the points of $\{\zeta_k\}$ and have the required orders, since

$$f(z) = (z - \zeta_k)^{\alpha_k} e^{\varphi_k(z)} \qquad (k = 1, 2, \ldots)$$

in a neighborhood of ζ_k.

By a function $f(z)$ *meromorphic on a domain* D we mean a function which can be written as a quotient

$$f(z) = \frac{g(z)}{h(z)}$$

of two functions $g(z)$ and $h(z)$ analytic on G, where $h(z) \not\equiv 0$. Thus the functions called meromorphic in the earlier volumes of this course are just those meromorphic in the finite plane [corresponding to the case where $g(z)$ and $h(z)$ are entire functions], and the above definition is the appropriate generalization to the case of an arbitrary domain D. To complete the picture, we now prove the generalization of Theorem II.10.8:

THEOREM 3.4. *A function $f(z)$ is meromorphic on a domain D if and only if its only singular points in D are poles.*

Proof. If $f(z)$ has a representation of the form $g(z)/h(z)$, where $g(z)$ and $h(z)$ are analytic on D, then $f(z)$ is analytic at every point where $h(z) \neq 0$. Suppose $h(z)$ has a zero of order $n \geqslant 1$ at a point $z_0 \in D$, and suppose z_0 is a zero of order $m \geqslant 0$ of $g(z)$. [If z_0 is not a zero of $g(z)$, we set $m = 0$.] Then $f(z)$ has a pole of order $n - m$ at z_0 if $n > m$ and a regular point at z_0 if $n \leqslant m$, where the regular point is a zero of order $m - n$ if $n < m$. Moreover, since the zeros of $h(z)$ cannot have a limit point in D [otherwise $h(z)$ would vanish identically], neither can the poles of $f(z)$. Thus the only singular points of $f(z)$ in D are poles, of which there are countably many (by Theorem I.17.3).

Conversely, let $f(z)$ be a function whose only singular points in D are poles. If $f(z)$ has no poles, then $f(z)$ is analytic on D and hence (trivially) meromorphic on D. Suppose $f(z)$ has a finite number of poles $\zeta_1, \ldots, \zeta_k$ in D, of orders $\alpha_1, \ldots, \alpha_k$, respectively. Then, forming the polynomial

$$P(z) = (z - \zeta_1)^{\alpha_1} \cdots (z - \zeta_k)^{\alpha_k},$$

we see that $g(z) = f(z)P(z)$ is analytic on D. In other words, $f(z)$ can be written as a quotient $g(z)/P(z)$ of two functions analytic on D, as required. Finally, suppose $f(z)$ has an infinite number of poles in D. Then no compact subset $F \subset D$ can contain more than a finite number of poles, since otherwise F would contain a limit point of poles, contrary to the assumption that $f(z)$ has no singular points in D other than poles. Therefore, by the same argument as used to prove Theorem I.17.3, $f(z)$ has countably many poles, which we write as a sequence $\{\zeta_k\}$ with its limit points on the boundary of D (there must be at least one such limit point), where ζ_k is of order α_k ($k = 1, 2, \ldots$). According to Theorem 3.3, there is a function $h(z)$ analytic on D whose zeros coincide with the poles of $f(z)$ and have the same orders. But then $g(z) = f(z)h(z)$ is analytic on D, i.e., we can write $f(z)$ as a quotient $g(z)/h(z)$ of two functions analytic on D, and the proof is complete.

Remark. Recalling the first paragraph of Sec. II.17, we note that the class of functions meromorphic on a domain G is the union of two disjoint subclasses, one consisting of the functions univalent on G, the other of the functions multivalent on G.

II. Runge's Theorem and Related Results

Turning now to the subject of approximation by polynomials, we prove

THEOREM 3.5 (*Runge's theorem*). *Let $\mathcal{O}$ be a union of (countably many) disjoint simply connected domains, none of which contains the point at infinity, and let $f(z)$ be locally analytic on $\mathcal{O}$. Then there exists a sequence of polynomials $\{P_n(z)\}$ converging uniformly to $f(z)$ inside $\mathcal{O}$.*

Proof. Let $\{\varepsilon_n\}$ be a sequence of positive numbers converging to zero. With the same notation as in the proof of Theorem 3.1, let $\{\mathcal{O}_n\}$ be an increasing sequence of open sets approximating the open set $\mathcal{O}$, and suppose that

$$|f(z) - S^{(n+1)}(z)| < \varepsilon_n$$

for all $z \in \bar{\mathcal{O}}_n$, where

$$S^{(n+1)}(z) = \frac{1}{2\pi i} \sum_{k=1}^{N^{(n+1)}} \frac{f(\zeta_k'^{(n+1)})}{\zeta_k'^{(n+1)} - z} (\zeta_k''^{(n+1)} - \zeta_k'^{(n+1)}).$$

Each component of $\mathcal{O}_n$ is a simply connected domain, since its boundary is a closed Jordan curve contained in one of the components of $\mathcal{O}$. Therefore, if $|z| < R_n$ is a disk containing $\bar{\mathcal{O}}_n$ but not some point $\tilde{\zeta}$, we can join $\tilde{\zeta}$ to every point of the boundary Γ_{n+1} of the set $\mathcal{O}_{n+1}$, in particular to every point $\zeta_k'^{(n+1)}$, by a Jordan arc which does not intersect $\bar{\mathcal{O}}_n$ (see Figure 3.2). Applying the lemma on p. 83 to every rational function

$$\frac{1}{2\pi i}\sum_{k=1}^{N^{(n+1)}} \frac{f(\zeta_k'^{(n+1)})}{\zeta_k'^{(n+1)} - z}(\zeta_k''^{(n+1)} - \zeta_k'^{(n+1)}) \qquad (k = 1, 2, \ldots, N^{(n+1)}),$$

and choosing

$$\varepsilon = \frac{\varepsilon_n}{N^{(n+1)}},$$

we construct a rational function

$$\tilde{S}_n(z) = \frac{\tilde{P}(z)}{(z - \tilde{\zeta})^{\tilde{n}}}$$

whose only pole is at $\tilde{\zeta}$,[7] such that

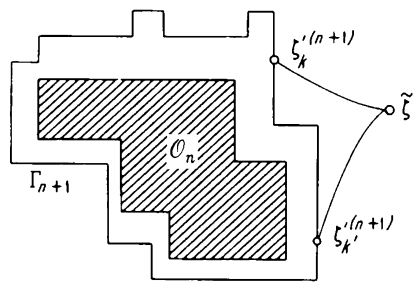


FIGURE 3.2

$$|\tilde{S}_n(z) - S^{(n+1)}(z)| < \varepsilon_n$$

and hence

$$|f(z) - \tilde{S}_n(z)| < 2\varepsilon_n$$

for all $z \in \bar{\mathcal{O}}_n$. But the function $\tilde{S}_n(z)$ is analytic on the disk $|z| < R_n$ containing $\bar{\mathcal{O}}_n$, and hence can be replaced by a power series on $\bar{\mathcal{O}}_n$. Choosing $P_n(z)$ to be the sum of a sufficiently large number of initial terms of this series, we find that

$$|f(z) - P_n(z)| < 3\varepsilon_n \qquad (z \in \bar{\mathcal{O}}_n),$$

i.e., the sequence $\{P_n(z)\}$ satisfies all the requirements of the theorem.

THEOREM 3.6. *Let $\mathcal{P}$ be a bounded perfect set,*[8] *and let G be the complement of $\mathcal{P}$ (relative to the extended plane). Suppose G is connected and hence a domain, and let $f(z)$ be locally analytic on $\mathcal{P}$. Then, given any $\varepsilon > 0$, there exists a polynomial $Q(z)$ such that*

$$|f(z) - Q(z)| < \varepsilon \qquad (z \in \mathcal{P}).$$

Proof. Since $f(z)$ is locally analytic on $\mathcal{P}$, for every point $z_0 \in \mathcal{P}$ there is a neighborhood $\mathcal{N}(z_0; \delta(z_0))$ such that $f(z)$ is the sum of a power series in the variable $z - z_0$ if $z \in \mathcal{N}(z_0; \delta(z_0)) \cap \mathcal{P}$. Consider the "half-size" neighborhoods $\mathcal{N}(z_0; \frac{1}{2}\delta(z_0))$, $z_0 \in \mathcal{P}$. According to the

[7] The degree of the polynomial $\tilde{P}(z)$ does not exceed $\tilde{n}$.

[8] A set is said to be *perfect* if it is closed and dense in itself.

Heine-Borel theorem (Theorem I.3.14), there exists a finite number of points $z_1, \ldots, z_n$ in $\mathscr{P}$ such that $\mathscr{P}$ is contained in the open set

$$\mathcal{O}' = \mathcal{N}(z_1; \tfrac{1}{2}\delta_1) \cup \cdots \cup \mathcal{N}(z_n; \tfrac{1}{2}\delta_n),$$

where $\delta_j = \delta(z_j)$, $j = 1, \ldots, n$. If $\mathcal{N}(z_j; \frac{1}{2}\delta_j) \cap \mathcal{N}(z_k; \frac{1}{2}\delta_k) \neq 0$, then

$$|z_j - z_k| < \max(\delta_j, \delta_k),$$

since

$$|z_j - z_k| < \tfrac{1}{2}(\delta_j + \delta_k).$$

Suppose, for example, that $\max(\delta_j, \delta_k) = \delta_j$. Then $\mathcal{N}(z_j; \delta_j)$ contains z_k and hence infinitely many points of $\mathscr{P}$. But the sums of the power series in $z - z_j$ and $z - z_k$ coincide on this set of points, and hence they coincide on $\mathcal{N}(z_j; \delta_j) \cap \mathcal{N}(z_k; \delta_k)$ as well. Therefore the series

$$\sum_{m=0}^{\infty} a_j^{(m)}(z - z_j)^m \qquad (j = 1, \ldots, n)$$

representing the function $f(z)$ in the neighborhoods $\mathcal{N}(z_j; \delta_j)$ define an analytic function on $\mathcal{O}'$ coinciding with $f(z)$ on $\mathscr{P}$. If $\mathcal{O}'$ is a union of disjoint simply connected domains, the proof is an immediate consequence of Runge's theorem, applied to the set $\mathcal{O}'$ and the function $f(z)$.

Thus suppose at least one component of $\mathcal{O}'$ is a multiply connected domain. Clearly, the boundary of $\mathcal{O}'$ consists of a finite number p of disjoint continua contained in G. Choosing points $\zeta_1, \ldots, \zeta_p$, one on each continuum, we join $\zeta_1, \ldots, \zeta_{p-1}$ to ζ_p by Jordan arcs $\gamma_1, \ldots, \gamma_{p-1}$ contained in G. Then

$$\mathcal{O} = \mathcal{O}' - \mathcal{O}' \cap (\gamma_1 \cup \cdots \cup \gamma_{p-1})$$

is an open set containing $\mathscr{P}$, all of whose components are disjoint simply

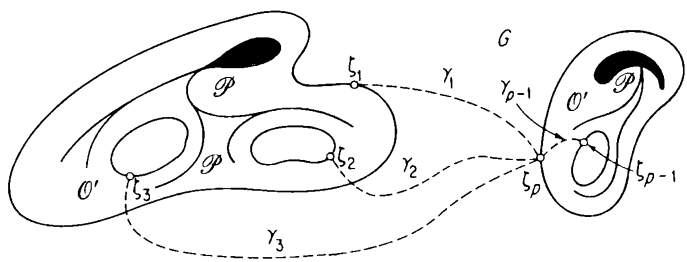


FIGURE 3.3

connected domains (see Figure 3.3). Since $\mathscr{P} \subset \mathcal{O} \subset \mathcal{O}'$, the proof is now completed by applying Runge's theorem to $\mathcal{O}$ and $f(z)$.

Using Theorem 3.6, we can prove the following stronger version of Runge's theorem:

THEOREM 3.7. *Let $\mathcal{O}$ be the union of a family of disjoint simply connected domains, none of which contains the point at infinity, and let $\mathcal{P}$ be a bounded perfect set, disjoint from $\mathcal{O}$, whose complement is connected (and hence a domain). Let $f(z)$ be locally analytic on $\mathcal{O}$, and let $\varphi(z)$ be locally analytic on $\mathcal{P}$. Then there exists a sequence of polynomials $\{P_n(z)\}$ converging to $f(z)$ on $\mathcal{O}$ and to $\varphi(z)$ on $\mathcal{P}$, and moreover the convergence is uniform on every set $F \cup \mathcal{P}$, where F is an arbitrary compact subset of $\mathcal{O}$.*

Proof. Let $\{\mathcal{O}_n\}$ be the increasing sequence of open sets approximating $\mathcal{O}$ constructed on p. 81, where, by hypothesis, every $\mathcal{O}_n$ is a union of disjoint simply connected domains. Clearly $\bar{\mathcal{O}}_n \cap \mathcal{P} = 0$ since $\bar{\mathcal{O}}_n \subset \mathcal{O}$. Therefore the function $g(z)$, equal to $f(z)$ on $\bar{\mathcal{O}}_n$ and $\varphi(z)$ on $\mathcal{P}$, is locally analytic on the perfect set $\bar{\mathcal{O}}_n \cup \mathcal{P}$, and moreover the complement of $\bar{\mathcal{O}}_n \cup \mathcal{P}$ is connected. Hence, according to Theorem 3.6, given any $\varepsilon_n > 0$, there exists a polynomial $P_n(z)$ satisfying the condition

$$|g(z) - P_n(z)| < \varepsilon_n \qquad (z \in \bar{\mathcal{O}}_n \cup \mathcal{P}).$$

Then the sequence $\{P_n(z)\}$ satisfies all the requirements of the theorem.

Example 1. Suppose we divide the z-plane into infinitely many squares with sides of unit length parallel to the coordinate axes, where, to be explicit, the origin is a vertex of one of the squares. Then we can construct a sequence of polynomials $\{P_n(z)\}$ converging in the whole plane, such that the limit of the sequence is one arbitrary entire function $g(z)$ at interior points of the squares and another arbitrary entire function $h(z)$ on the sides of the squares. To see this, consider the set

$$-n \leqslant x \leqslant n, \qquad -n \leqslant y \leqslant n, \tag{3.9}$$

where n is an arbitrary positive integer, and replace each of the unit squares contained in (3.9) by a set of five closed squares and four closed rectangles, as shown in Figure 3.4, which indicates the dimensions of the squares and rectangles and of the spaces between them.[9] Let F_n be the union of all the closed squares and rectangles associated with the large square (3.9). Then

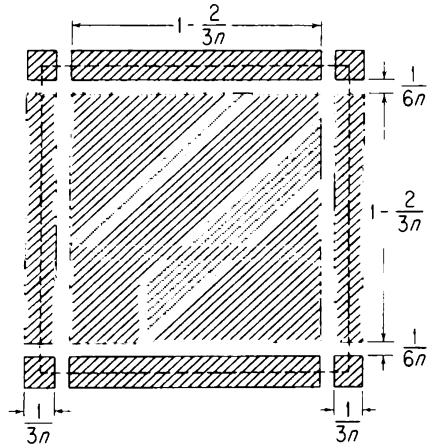


FIGURE 3.4

[9] Note that some rectangles are "shared" by two original squares and some small squares by four original squares. The spaces between the squares and rectangles are crucial to the construction (why?).

F_n is a bounded perfect set whose complement is connected. As n increases, the small squares contract, the inner squares concentric with the original squares expand, and the rectangles lengthen, at the same time as the total number of squares and rectangles increases. Clearly, for every point z_0, there is an integer $N(z_0) > 0$ such that $z_0 \in F_n$ for all $n > N(z_0)$. In fact, a point which is a vertex of one of the original squares will eventually belong to and stay in one of the small squares, a point which belongs to a side of one of the original squares but is not a vertex will eventually belong to and stay in one of the rectangles, while a point which is an interior point of one of the original squares will eventually belong to and stay in one of the squares concentric with the original squares. Suppose we define a locally analytic function $f_n(z)$ on F_n by setting $f_n(z) = g(z)$ on the squares concentric with the original squares and $f_n(z) = h(z)$ on the eight squares and rectangles bordering each original square. Then, according to Theorem 3.6, there exists a polynomial $P_n(z)$ such that

$$|f_n(z) - P_n(z)| < \frac{1}{n} \qquad (z \in F_n),$$

i.e., the sequence $\{P_n(z)\}$ satisfies all the requirements of our problem.

Example 2. Using Theorem 3.7, we can construct an entire function $f(z)$ with the following remarkable property: Given any bounded simply connected domain G and any function $\varphi(z)$ analytic on G, there is a sequence of positive integers $\{n_m\}$ such that the sequence $\{f(z + n_m)\}$ converges uniformly to $\varphi(z)$ inside G. In other words, $f(z)$ is a "universal function" in the sense that it can be used to approximate any analytic function on any bounded simply connected domain provided only that the points of the z-plane are subjected to appropriate shifts n_m $(m = 1, 2, \ldots)$.

We begin the construction by considering the set of all polynomials with coefficients whose real and imaginary parts are rational numbers. Since the set of all such polynomials is countable (cf. Problem I.3.2), they can be ordered in a single sequence $\{\Pi_n(z)\}$. Now consider the sequence of disjoint closed disks

$$\bar{K}_n: |z - n^3| \leqslant n \qquad (n = 1, 2, \ldots),$$

observing that $\bar{K}_n$ is contained in the annulus

$$(n-1)^3 + (n-1)^2 + (n-1) + 1 < |z| < n^3 + n^2 + n + 1$$

if $n > 1$. Let $P_1(z)$ be an arbitrary polynomial, and suppose $P_{n-1}(z)$, $n > 1$ has already been determined. Then, using Theorem 3.7, we determine the next polynomial $P_n(z)$ so as to satisfy the following two inequalities:

$$|P_n(z) - P_{n-1}(z)| < \frac{1}{2^n} \text{ if } |z| < (n-1)^3 + (n-1)^2 + (n-1) + 1,$$

$$|P_n(z) - \Pi_n(z - n^3)| < \frac{1}{2^n} \text{ if } |z - n^3| \leqslant n.$$

Obviously, the series

$$P_1(z) + \sum_{n=2}^{\infty} [P_n(z) - P_{n-1}(z)] \equiv f(z)$$

is uniformly convergent on any disk $|z| < R$, since the absolute values of its terms, starting from some value of n, are less than the corresponding terms of the convergent numerical series

$$\sum_{n=1}^{\infty} \frac{1}{2^n}.$$

Therefore $f(z)$ is analytic in the finite plane, i.e., entire. Moreover, on every disk $\bar{K}_n$, $f(z)$ satisfies the inequality

$$\begin{aligned} |f(z) - \Pi_n(z - n^3)| &\leqslant |f(z) - P_n(z)| + |P_n(z) - \Pi_n(z - n^3)| \\ &\leqslant \sum_{k=n+1}^{\infty} |P_k(z) - P_{k-1}(z)| + |P_n(z) - \Pi_n(z - n^3)| \\ &< \sum_{k=n+1}^{\infty} \frac{1}{2^k} + \frac{1}{2^n} = \frac{1}{2^{n-1}}. \end{aligned}$$

(Here we use the fact that $\bar{K}_n$ is contained in the disk $|z| < n^3 + n^2 + n + 1$.) It follows that $f(z)$ satisfies the inequality

$$|f(z + n^3) - \Pi_n(z)| < \frac{1}{2^{n-1}}$$

on every disk $|z| \leqslant n$.

Now let G be any bounded simply connected domain, and let $\varphi(z)$ be analytic on G. According to Runge's theorem, there exists a sequence of polynomials $\{p_m(z)\}$ converging uniformly to $\varphi(z)$ inside G. But for every polynomial $p_m(z)$, we can find infinitely many polynomials $\Pi(z)$, with coefficients whose real and imaginary parts are rational numbers, such that

$$|p_m(z) - \Pi(z)| < \frac{1}{m} \qquad (z \in \bar{G}).$$

These polynomials belong to the sequence $\{\Pi_n(z)\}$, and in fact have arbitrarily large indices in this sequence (since there are infinitely many of them). Choosing $\Pi(z) = \Pi_{n_m}(z)$ such that $n_m > m$ (then $m \to \infty$ implies $n_m \to \infty$), we have

$$|p_m(z) - \Pi_{n_m}(z)| < \frac{1}{m} \qquad (z \in \bar{G}).$$

The sequence of polynomials $\{\Pi_{n_m}(z)\}$ obviously converges uniformly to $\varphi(z)$ inside G, and moreover, if n_m is so large that G is contained in the disk $|z| < n_m$, then

$$|\Pi_{n_m}(z) - f(z + n_m^3)| < \frac{1}{2^{n_m - 1}} \qquad (z \in \bar{G}),$$

by the construction of the entire function $f(z)$. Therefore the sequence $\{f(z + n_m^3)\}$ also converges uniformly to $\varphi(z)$ inside G, i.e., $f(z)$ is the required "universal function."

Example 3. Finally we use Runge's theorem to construct a function $f(z)$ which is analytic on the unit disk $K: |z| < 1$ but approaches no limit as z approaches an arbitrary point ζ of the unit circle $|z| = 1$ along the radius joining the origin to ζ. (Such a function is said to have no radial boundary values.) Let

$$0 < r_0 < r_1 < \cdots < r_n < \cdots,$$

where

$$\lim_{n\to\infty} r_n = 1,$$

and define the open sets

$$B_0: |z| < r_0,$$

$$B_1: r_1 < |z| < r_2, \quad |\arg z| < \frac{3\pi}{4} \quad \text{or} \quad r_3 < |z| < r_4, \quad |\arg z| > \frac{\pi}{4},$$

$$B_2: r_5 < |z| < r_6, \quad |\arg z| < \frac{3\pi}{4} \quad \text{or} \quad r_7 < |z| < r_8, \quad |\arg z| > \frac{\pi}{4},$$

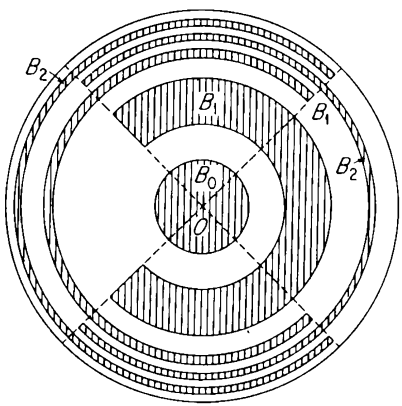


FIGURE 3.5

and so on, where in the general case,

$$B_n: r_{4n-3} < |z| < r_{4n-2}, \quad |\arg z| < \frac{3\pi}{4}$$

or

$$r_{4n-1} < |z| < r_{4n}, \quad |\arg z| > \frac{\pi}{4} \qquad (n = 1, 2, \ldots)$$

(see Figure 3.5).

Next we construct a sequence of polynomials $\{P_n(z)\}$ by starting with $P_0(z) \equiv 0$ and then requiring that $P_{2n-1}(z)$ and $P_{2n}(z)$ satisfy the following inequalities [under the assumption that $P_0(z), \ldots, P_{2n-2}(z)$ have already been constructed]:

$$|P_{2n-1}(z) - P_{2n-2}(z)| < \frac{1}{2^{2n-1}} \quad \text{if} \quad |z| < r_{8n-8},$$

$$|P_{2n-1}(z) - 1| < \frac{1}{2^{2n}} \quad \text{if} \quad z \in B_{2n-1},$$

and

$$|P_{2n}(z) - P_{2n-1}(z)| < \frac{1}{2^{2n}} \quad \text{if} \quad |z| < r_{8n-4},$$

$$|P_{2n}(z)| < \frac{1}{2^{2n}} \quad \text{if} \quad z \in B_{2n}.$$

Given any compact subset $F \subset K$, F is contained in the disk $|z| < r_{4n}$ for all sufficiently large $n > N(F)$. Therefore

$$|P_{n+1}(z) - P_n(z)| < \frac{1}{2^{n+1}} \qquad (z \in F)$$

if $n > N(F)$, i.e., $\{P_n(z)\}$ is uniformly convergent on every compact subset of K, and hence represents an analytic function $f(z)$ on K. Moreover

$$\begin{aligned} |f(z) - 1| &\leqslant |f(z) - P_{2n-1}(z)| + |P_{2n-1}(z) - 1| \\ &= \left| \sum_{j=2n-1}^{\infty} [P_{j+1}(z) - P_j(z)] \right| + |P_{2n-1}(z) - 1| \\ &< \sum_{j=2n-1}^{\infty} \frac{1}{2^{j+1}} + \frac{1}{2^{2n-1}} = \frac{1}{2^{2n-2}} \end{aligned}$$

for all $z \in B_{2n-1}$, while

$$|f(z)| \leqslant \left| \sum_{j=2n}^{\infty} [P_{j+1}(z) - P_j(z)] \right| + |P_{2n}(z)| < \sum_{j=2n}^{\infty} \frac{1}{2^{j+1}} + \frac{1}{2^{2n}} = \frac{1}{2^{2n-1}}$$

for all $z \in B_{2n}$. It follows from these inequalities and the relative positions of the sets B_n that $f(z)$ has no radial boundary values.

We now turn to the case where $\mathcal{O}$ is not a union of disjoint simply connected domains. Thus again let $\mathcal{O}$ be an open set which does not contain the point at infinity, but this time suppose $\mathcal{O}$ has at least one multiply connected component. Then if $f(z)$ is locally analytic on $\mathcal{O}$, it is in general no longer possible to find a sequence of polynomials $\{P_n(z)\}$ converging uniformly to $f(z)$ inside $\mathcal{O}$. In fact, let Γ be any closed Jordan curve contained in $\mathcal{O}$, with interior $I(\Gamma)$. Then, since Γ is a compact subset of $\mathcal{O}$, if $\{P_n(z)\}$ converges uniformly inside $\mathcal{O}$, $\{P_n(z)\}$ converges uniformly on Γ and hence on $I(\Gamma)$ (cf. Theorem I.15.7). Therefore $\{P_n(z)\}$ converges uniformly to a locally analytic function inside the set

$$\mathcal{O}' = \bigcup_{\Gamma \subset \mathcal{O}} I(\Gamma)$$

(cf. Problem 3.7, part c), equal to the union of the interiors of all closed Jordan curves contained in $\mathcal{O}$.[10] In other words, we can find a sequence $\{P_n(z)\}$ converging uniformly to $f(z)$ inside $\mathcal{O}$ only in the case where $f(z)$ is the restriction to $\mathcal{O}$ of a locally analytic function on $\mathcal{O}'$, i.e., where there exists a locally analytic function $\varphi(z)$ on $\mathcal{O}'$ such that $\varphi(z) \equiv f(z)$ on $\mathcal{O}$.

Example. The function

$$f(z) = \frac{1}{z}$$

is locally analytic on the open set $\mathcal{O}$ which is the union of the two domains

[10] $\mathcal{O}'$ is necessarily a union of *disjoint* simply connected domains (cf. Problem 3.7, part b).

$0 < |z| < 1$ and $|z - 2| < 1$, but is not the restriction to $\mathcal{O}$ of any function locally analytic on the set $\mathcal{O}'$, which in this case is the union of the two domains $|z| < 1$ and $|z - 2| < 1$.

As we have just seen, if $\mathcal{O}$ is an open set with at least one multiply connected component and if $f(z)$ is locally analytic on $\mathcal{O}$, then in general there is no sequence of polynomials $\{P_n(z)\}$ converging uniformly to $f(z)$ inside $\mathcal{O}$. However, we can still use Theorem 3.1 to find a sequence of rational functions $\{R_n(z)\}$ (which do not reduce to polynomials) converging uniformly to $f(z)$ inside $\mathcal{O}$ ($\mathcal{O}$ can now contain the point at infinity). It will be recalled that because of the way $R_n(z)$ is constructed, all its poles lie on the boundary Γ_{n+1} of the open set $\mathcal{O}_{n+1}$, where $\bar{\mathcal{O}}_n \subset \mathcal{O}_{n+1} \subset \mathcal{O}$. The complement of $\bar{\mathcal{O}}_n$ relative to the extended plane consists of a finite number of domains $D_1^{(n)}, \ldots, D_{\nu_n}^{(n)}$, each of which contains at least one boundary point or exterior point of $\mathcal{O}$, and hence at least one of the closed polygonal Jordan curves making up Γ_{n+1}. Suppose that in each domain $D_j^{(n)}$ we choose a point $\alpha_j^{(n)}$ which is a boundary point or exterior point of $\mathcal{O}$. Then, by the lemma on p. 83 (used to prove Runge's theorem), the function $R_n(z)$ can be replaced by another rational function $\tilde{R}_n(z)$ whose poles coincide with the points $\alpha_1^{(n)}, \ldots, \alpha_{\nu_n}^{(n)}$ and which satisfies the condition

$$|\tilde{R}_n(z) - R_n(z)| < \frac{1}{n} \qquad (z \in \bar{\mathcal{O}}_n).$$

It follows that there exists a sequence of rational functions $\{\tilde{R}_n(z)\}$ converging uniformly to $f(z)$ inside $\mathcal{O}$, where all the poles of the functions $\tilde{R}_n(z)$ are either boundary points or exterior points of $\mathcal{O}$. Moreover, the choice of these poles is arbitrary to within the limits of the construction just described.

Example. Suppose $\mathcal{O}$ is a finitely connected domain whose boundary consists of $p \geqslant 2$ disjoint closed Jordan curves $\gamma_1, \ldots, \gamma_p$. Then each curve γ_j is the boundary of a domain disjoint from $\mathcal{O}$. Choosing a point α_j ($j = 1, \ldots, p$) in each of these domains, we can require that the poles of $\tilde{R}_n(z)$ coincide with the points $\alpha_1, \ldots, \alpha_p$. In particular, if $\mathcal{O}$ is the annulus

$$0 < r < |z - z_0| < R < \infty,$$

then $p = 2$, and choosing $\alpha_1 = z_0$ and $\alpha_2 = \infty$, we find that the function $\tilde{R}_n(z)$ has the form

$$\begin{aligned} \tilde{R}_n(z) &= \frac{A_0^{(n)} + A_1^{(n)}(z - z_0) + \cdots + A_{\lambda_n}^{(n)}(z - z_0)^{\lambda_n}}{(z - z_0)^{\mu_n}} \\ &= \sum_{j=-\mu_n}^{\lambda_n - \mu_n} A_j^{(n)}(z - z_0)^j \end{aligned}$$

where $\lambda_n > \mu_n$. Therefore every function analytic on the annulus has the representation

$$f(z) = \lim_{n \to \infty} \tilde{R}_n(z) = \lim_{n \to \infty} \sum_{j=-\mu_n}^{\lambda_n - \mu_n} A_j^{(n)}(z - z_0)^j. \tag{3.10}$$

However, as we know, Laurent's theorem (Theorem II.1.3) allows us to say more in this case. In fact, the coefficients $A_j^{(n)}$ need not depend on n, so that (3.10) becomes

$$f(z) = \sum_{j=-\infty}^{\infty} A_j(z - z_0)^j.$$

If we drop the requirement of uniform convergence, the problem of polynomial approximation of a locally analytic function on an open set becomes more tractable:

THEOREM 3.8 (*Montel's theorem*). *Let $\mathcal{O}$ be an arbitrary open set which does not contain the point at infinity, and let $f(z)$ be locally analytic on $\mathcal{O}$. Then there exists a sequence of polynomials $\{P_n(z)\}$ converging to $f(z)$ on $\mathcal{O}$.*

Proof. Let $\{\mathcal{O}_n\}$ be the increasing sequence of open sets approximating $\mathcal{O}$ constructed on p. 81. Suppose we represent $\mathcal{O}_n$ in the form

$$\mathcal{O}_n = \mathcal{O}_1 \cup (\mathcal{O}_2 - \mathcal{O}_1) \cup \cdots \cup (\mathcal{O}_n - \mathcal{O}_{n-1}) = \Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_n.$$

By construction, each set Ω_j is made up of equal squares with sides of length 3^{-j} ($j = 1, \ldots, n$). Suppose we replace each of these squares by five squares and four rectangles, just as was done in Example 1, p. 91 (see Figure 2.4), except that now the squares have side length 3^{-j} instead of 1. Let F_n be the union of all the squares and rectangles associated with $\Omega_1, \Omega_2, \ldots, \Omega_n$. Then F_n is a bounded perfect set whose complement is connected. Finally, let $\{\varepsilon_n\}$ be a sequence of positive numbers converging to zero. Since $F_n \subset \mathcal{O}$ (in fact $F_n \subset \mathcal{O}_{n+1}$), then, given any ε_n, we can use Theorem 3.6 to find a polynomial $P_n(z)$ such that

$$|P_n(z) - f(z)| < \varepsilon_n.$$

To prove that the sequence $\{P_n(z)\}$ converges to $f(z)$ on $\mathcal{O}$, we need only show that for every point $z_0 \in \mathcal{O}$, there is an integer $N(z_0) > 0$ such that $z_0 \in F_n$ for all $n > N(z_0)$, since $|P_n(z_0) - f(z_0)| < \varepsilon_n$ for all $n > N(z_0)$, i.e.,

$$\lim_{n\to\infty} P_n(z_0) = f(z_0).$$

But this follows by exactly the same argument as given in the cited example, since every point $z_0 \in \mathcal{O}$ belongs to some set Ω_j and hence to the sides or the interior of one of the "original squares" making up Ω_j.[11]

12. Approximation on Closed Domains

First we consider polynomial approximation on a closed *Jordan* domain:

[11] Note that the sizes of the five squares and four rectangles contributed to F_n by each original square in Ω_j depends on n.

THEOREM 3.9 (*Walsh's theorem*). *Let G be the interior of a closed Jordan curve Γ, and suppose $\varphi(z)$ is analytic on G and continuous on $\bar{G}$. Then, given any $\varepsilon > 0$, there exists a polynomial $P(z)$ such that*

$$|\varphi(z) - P(z)| < \varepsilon \qquad (z \in \bar{G}). \tag{3.11}$$

Proof. Let $w = f(z)$ be the function mapping the domain G conformally onto the disk $K: |w| < 1$ and satisfying the conditions $f(z_0) = 0$, $f'(z_0) > 0$ where $z_0 \in G$. Since the function $z = f^{-1}(w)$ is continuous on $\bar{K}$ (when extended to the unit circle as in Theorem 2.21), the function $\varphi^*(w) = \varphi[f^{-1}(w)]$ is continuous on $\bar{K}$, as well as analytic on K. Therefore $\varphi^*(w)$ is uniformly continuous on $\bar{K}$, so that given any $\varepsilon > 0$, we have

$$|\varphi^*(e^{i\theta}) - \varphi^*(re^{i\theta})| < \frac{\varepsilon}{4}$$

for all $r \geqslant \rho(\varepsilon)$. On the other hand, the uniform convergence of the Taylor series of $\varphi^*(w)$ inside K implies

$$\left|\varphi^*[\rho(\varepsilon)e^{i\theta}] - \sum_{k=0}^{n} \frac{\varphi^{*(k)}(0)}{k!} [\rho(\varepsilon)]^k e^{ik\theta}\right| < \frac{\varepsilon}{4}$$

for all $0 \leqslant \theta \leqslant 2\pi$ and $n \geqslant N(\varepsilon)$. Writing

$$w = e^{i\theta}, \qquad a_k = \frac{\varphi^{*(k)}(0)}{k!} [\rho(\varepsilon)]^k,$$

we find that

$$\left|\varphi^*(w) - \sum_{k=0}^{N(\varepsilon)} a_k w^k\right| < \frac{\varepsilon}{2}$$

for all w on the unit circle. Using the mapping $z = f^{-1}(w)$ to return to the domain G, we obtain

$$\left|\varphi(z) - \sum_{k=0}^{N(\varepsilon)} a_k [f(z)]^k\right| < \frac{\varepsilon}{2} \tag{3.12}$$

for all $z \in \Gamma$. According to the maximum modulus principle, this inequality also holds for all $z \in \bar{G}$.

Next we use Theorem 2.2 to construct a decreasing sequence of bounded simply connected domains converging to G as its kernel. Let $w_n = f_n(z)$ be the function mapping G_n conformally onto K and satisfying the conditions $f_n(z_0) = 0, f_n'(z_0) > 0$. Then, according to Theorem 2.26, the sequence $\{f_n(z)\}$ converges uniformly to $f(z)$ on $\bar{G}$. For each $n = 1, 2, \ldots$, we now use Runge's theorem (Theorem 3.5) to construct a polynomial $p_n(z)$ such that

$$|f_n(z) - p_n(z)| < \frac{1}{n} \qquad (z \in \bar{G}),$$

relying on the fact that $f_n(z)$ is analytic on the simply connected domain G_n containing $\bar{G}$. Obviously, the sequence $\{p_n(z)\}$ also converges uniformly to $f(z)$ on $\bar{G}$, and hence the sequence

$$\left\{\sum_{k=0}^{N(\varepsilon)} a_k[p_n(z)]^k\right\}$$

converges uniformly to

$$\sum_{k=0}^{N(\varepsilon)} a_k[f(z)]^k$$

on $\bar{G}$ (as $n \to \infty$). Let the integer $\nu(\varepsilon) > 0$ be such that

$$\left|\sum_{k=0}^{N(\varepsilon)} a_k[f(z)]^k - \sum_{k=0}^{N(\varepsilon)} a_k[p_n(z)]^k\right| < \frac{\varepsilon}{2} \tag{3.13}$$

for all $n \geqslant \nu(\varepsilon)$. Comparing (3.12) and (3.13), and denoting the polynomial

$$\sum_{k=0}^{N(\varepsilon)} a_k[p_{\nu(\varepsilon)}(z)]^k$$

by $P(z)$, we obtain the desired result (3.11).

As an important consequence of Walsh's theorem, we have

THEOREM 3.10 (*Generalization of Cauchy's integral theorem*).[12] *If G is the interior of a closed rectifiable Jordan curve Γ, and if $\varphi(z)$ is analytic on G and continuous on $\bar{G}$, then*

$$\int_\Gamma \varphi(z)\, dz = 0. \tag{3.14}$$

Proof. According to Theorem 3.9, given any $\varepsilon > 0$, there is a polynomial $P(z)$ such that

$$|\varphi(z) - P(z)| < \frac{\varepsilon}{l} \qquad (z \in \bar{G}),$$

where l is the length of Γ. Therefore

$$\left|\int_\Gamma \varphi(z)\, dz\right| = \left|\int_\Gamma [\varphi(z) - P(z)]\, dz\right| < \frac{\varepsilon}{l}\, l = \varepsilon,$$

which implies (3.14), since ε is arbitrary.

COROLLARY (*Generalization of Cauchy's integral formula*). *If G, Γ and $\varphi(z)$ are the same as in Theorem* 3.10, *then*

$$\frac{1}{2\pi i}\int_\Gamma \frac{\varphi(\zeta)}{\zeta - z}\, d\zeta = \varphi(z) \qquad \text{if} \quad z \in G,$$

$$\frac{1}{2\pi i}\int_\Gamma \frac{\varphi(\zeta)}{\zeta - z}\, d\zeta = 0 \qquad \text{if} \quad z \in (\bar{G})^c.$$

---

[12] Theorem 3.10 has already been proved in the special case where G is starlike (see Theorem I.13.3 and footnote 8, p. I.270).

Proof. Compare with the proof of the usual form of Cauchy's integral formula (Theorem I.14.1).

Walsh's theorem is just a special case of the following more general

THEOREM 3.11 (*Keldysh's theorem*). *Every function $\varphi(z)$ analytic on a domain G and continuous on $\overline{G}$ can be approximated arbitrarily closely on $\overline{G}$ by a polynomial if and only if $(\overline{G})^c$ consists of only one domain G_∞ containing the point at infinity.*

Remark 1. Of course, all Jordan domains are of the type required by Keldysh's theorem, but there are other such domains, e.g., the domain shown in Figure 2.18, whose boundary is not a Jordan curve.

Remark 2. The proof of the necessity is almost obvious. In fact, suppose the conditions of Keldysh's theorem are not satisfied. Then there exists a domain $G_1 \neq G_\infty$ which is a component of $(\overline{G})^c$. If $z_0 \in G_1$, then the function

$$\frac{1}{z - z_0} \tag{3.15}$$

is analytic on G and continuous on $\overline{G}$. If there were a sequence of polynomials $\{P_n(z)\}$ converging uniformly to (3.15) on $\overline{G}$, then $\{P_n(z)\}$ would converge uniformly on the boundary of G_∞, which is contained in the boundary of $\overline{G}$. Therefore $\{P_n(z)\}$ would converge uniformly to an analytic function $\Phi(z)$ on the domain $(\overline{G}_\infty)^c$ containing both G and G_1. But this is impossible, since $\Phi(z)$ coincides with (3.15) on G and hence coincides with (3.15) on G_1 as well, contrary to the analyticity of $\Phi(z)$.

Thus the whole difficulty of the proof is in the sufficiency part, for which we refer the reader to the original paper,[13] noting only that the proof requires the use of a deep result due to Lavrentev (see Theorem 3.13 below).

13. Approximation on Continua

First we consider rational approximation on a closed Jordan curve:

THEOREM 3.12. *Let Γ be a closed Jordan curve with interior $I(\Gamma)$, and let $\varphi(z)$ be continuous on Γ. Then, given any $z_0 \in I(\Gamma)$ and $\varepsilon > 0$, there exists a rational function*

$$S(z) = \sum_{k=-\nu}^{\nu} a_n(z - z_0)^n \tag{3.16}$$

such that

$$|\varphi(z) - S(z)| < \varepsilon \qquad (z \in \Gamma). \tag{3.17}$$

[13] M. V. Keldysh, *On the representation of functions of a complex variable by series of polynomials on closed domains* (in Russian), Mat. Sb., **16**, 249 (1945).

Proof. Suppose $w = f(z)$ maps $I(\Gamma)$ conformally onto the unit disk $K: |w| < 1$. Since $z = f^{-1}(w)$ is continuous on $\bar{K}$, the function $\varphi^*(w) = \varphi[f^{-1}(w)]$ is continuous on the unit circle, where $\varphi^*(w)$ can be written in the form

$$\varphi^*(w) = \varphi^*(e^{i\theta}) = \chi(\theta) + i\psi(\theta),$$

in terms of two continuous real-valued functions $\chi(\theta)$ and $\psi(\theta)$, of period 2π since $\varphi^*(w)$ is single-valued. According to Weierstrass' approximation theorem (Theorem II.5.8, Corollary 2),[14] given any $\varepsilon > 0$, we can find trigonometric polynomials $\sigma(\theta)$ and $\tau(\theta)$ such that

$$|\chi(\theta) - \sigma(\theta)| < \frac{\varepsilon}{4}, \quad |\psi(\theta) - \tau(\theta)| < \frac{\varepsilon}{4} \qquad (0 \leqslant \theta \leqslant 2\pi).$$

Introducing extra coefficients equal to zero (if necessary), we can write $\sigma(\theta)$ and $\tau(\theta)$ as trigonometric polynomials of the same degree N:

$$\sigma(\theta) = a_0 + \sum_{n=1}^{N} (a_n \cos n\theta + b_n \sin n\theta) = \sum_{n=-N}^{N} \alpha_n e^{in\theta},$$

$$\tau(\theta) = c_0 + \sum_{n=1}^{N} (c_n \cos n\theta + d_n \sin n\theta) = \sum_{n=-N}^{N} \beta_n e^{in\theta}.$$

Replacing $e^{i\theta}$ by w, we obtain

$$|\chi(\theta) + i\psi(\theta) - [\sigma(\theta) + i\tau(\theta)]| = \left|\varphi^*(w) - \sum_{n=-N}^{N} \gamma_n w^n\right| < \frac{\varepsilon}{2} \qquad (\gamma_n = \alpha_n + i\beta_n),$$

or using the transformation $w = f(z)$ to return to the z-plane,

$$\left|\varphi(z) - \sum_{n=-N}^{N} \gamma_n [f(z)]^n\right| < \frac{\varepsilon}{2} \qquad (z \in \Gamma).$$

According to Walsh's theorem, we can approximate $f(z)$ on Γ to within $\varepsilon/4$ by a polynomial $P(z)$, where $P(z)$ does not vanish on Γ (since $|f(z)| = 1 \neq 0$ on Γ). Thus there is a rational function

$$R(z) = \sum_{n=-N}^{N} \gamma_n [P(z)]^n,$$

with no poles on Γ, such that

$$|\varphi(z) - R(z)| < \frac{3\varepsilon}{4} \qquad (z \in \Gamma).$$

Given any point $z_0 \in I(\Gamma)$, we can find a doubly connected domain D containing Γ and no poles of $R(z)$ such that the boundary of D consists

---

[14] For another proof of Weierstrass' approximation theorem, see e.g., G. P. Tolstov, *Fourier Series* (translated by R. A. Silverman), Prentice-Hall, Inc., Englewood Cliffs, N. J. (1962), Chap. 5, Sec. 1.

of two disjoint Jordan curves, one inside the other, and z_0 belongs to the interior of the "inside" curve (recall Problem 2.13). Then, by the same argument as in the example on p. 96, we can approximate $R(z)$ on Γ to within $\varepsilon/4$ by another rational function $S(z)$ whose poles are z_0 and ∞. In other words, there is a rational function of the form (3.16) which satisfies the inequality (3.17). This completes the proof.

COROLLARY. *Let Γ be a closed Jordan curve, let $\varphi(z)$ be continuous on Γ, and let $\gamma \subset \Gamma$ be any arc with distinct end points. Then, given any $\varepsilon > 0$, there exists a polynomial*

$$P(z) = \sum_{n=0}^{\mu} a_n z^n \tag{3.18}$$

such that

$$|\varphi(z) - P(z)| < \varepsilon \qquad (z \in \gamma). \tag{3.19}$$

Proof. Applying Theorem 3.6 to the perfect set γ and the locally analytic function (3.16), we can find a polynomial (3.18) such that

$$|S(z) - P(z)| < \varepsilon \qquad (z \in \gamma). \tag{3.20}$$

But then (3.20) and (3.17) immediately imply (3.19), after changing 2ε to ε.

Remark. In the corollary just proved, it is important that γ have distinct end points. In fact, a function $\varphi(z)$ analytic on a closed Jordan curve Γ must satisfy extra conditions if $\varphi(z)$ is to be approximated on Γ to arbitrary accuracy by polynomials. For example, suppose Γ is a closed rectifiable Jordan curve, and let $\varphi(z)$ be a continuous function on Γ such that

$$\varphi(z) = \lim_{n \to \infty} P_n(z),$$

where $\{P_n(z)\}$ is a sequence of polynomials converging uniformly on Γ. Then obviously we must have

$$\int_\Gamma \varphi(z)\, dz = \lim_{n \to \infty} \int_\Gamma P_n(z)\, dz = 0,$$

which is one of the extra conditions in question. But there are others as well, since, given any integer $m > 0$, we have

$$\varphi(z) z^m = \lim_{n \to \infty} P_n(z) z^m,$$

and hence

$$\int_\Gamma \varphi(z) z^m\, dz = 0 \qquad (m = 1, 2, \ldots).$$

An arbitrary continuous function on Γ cannot be expected to satisfy all these conditions.

Theorem 3.12 generalizes Weierstrass' approximation theorem, but is itself only a special case of the following more general

THEOREM 3.13 (*Lavrentev's theorem*). *Every function $\varphi(z)$ continuous on a continuum F can be approximated arbitrarily closely on F by a polynomial if and only if F has no interior points and F^c is connected.*[15]

Remark. The proof of the necessity is very easy. In fact, if F has interior points, then the limit of a uniformly convergent sequence of polynomials on F must be analytic in a neighborhood of every such point. But then a continuous, nowhere analytic function like $\varphi(z) = \bar{z}$ cannot be approximated arbitrarily closely on F by polynomials. Thus suppose F has no interior points but separates the plane. Then at least one component of F^c, say G, is bounded. Let Γ be the boundary of G. If $z_0 \in G$, then the function

$$\frac{1}{z - z_0} \tag{3.21}$$

is obviously continuous on F. If there were a sequence of polynomials $\{P_n(z)\}$ converging uniformly to (3.21) on F, then $\{P_n(z)\}$ would converge uniformly on the boundary Γ, which is contained in F. Therefore $\{P_n(z)\}$ would converge uniformly on the closed domain $\bar{G}$ to a function $\Phi(z)$ which is analytic on G and continuous on $\bar{G}$, and coincides with (3.21) on Γ. Then the function

$$\Phi(z) - \frac{1}{z - z_0},$$

with a simple pole at the point $z_0 \in G$, would vanish on Γ, and hence the function

$$\Psi(z) = (z - z_0)\left[\Phi(z) - \frac{1}{z - z_0}\right] = (z - z_0)\Phi(z) - 1$$

would be analytic on G and continuous on $\bar{G}$, and would vanish on Γ. But then $\Psi(z) \equiv 0$, by the maximum modulus principle, which is impossible, since $\Psi(z_0) = -1$. This contradiction proves the necessity of the conditions figuring in Theorem 3.13.

Thus the whole difficulty of the proof is in the sufficiency part, for which we refer the reader to the original paper.[16] However, it should be noted that Theorems 3.11 and 3.13 are both special cases of even more general results recently proved by Mergelyan.[17]

---

[15] See Problem I.4.14. In other words, F must be a "linear continuum which does not separate the plane." A Jordan arc with distinct end points is the simplest example of such a continuum.

[16] M. A. Lavrentev, *Zur Theorie der konformen Abbildungen*, Travaux de l'Institut Physico-mathématique Stekloff (Leningrad), **5**, 159 (1934).

[17] See p. 367 of the book by Walsh (cited in the Bibliography on pp. 351–353).

14. Faber Polynomials

We now develop a theory which extends the theory of power series to domains more general than a disk, in a sense that will soon be apparent:

THEOREM 3.14. *Let F be a continuum containing more than one point, and let G_∞ be the component of F^c containing the point at infinity.*[18] *Then there is a function $w = \Phi(z)$ which maps G_∞ conformally onto a domain of the form $|w| > \rho > 0$ (the exterior of a circle) and satisfies the conditions*

$$\Phi(\infty) = \infty, \qquad \lim_{z \to \infty} \frac{\Phi(z)}{z} = 1. \tag{3.22}$$

Proof. Let the domain G be the image of G_∞ under the mapping

$$z_1 = \frac{1}{z - z_0} \qquad (z_0 \in F)$$

carrying the point $z = \infty$ into the point $z_1 = 0$, and let the function $w_1 = \varphi(z_1)$ map G onto the disk $|w_1| < \rho$, while satisfying the conditions $\varphi(0) = 0$ and

$$\varphi'(0) = 1, \qquad \text{i.e.,} \quad \lim_{z_1 \to 0} \frac{w_1}{z_1} = 1.$$

Then the function

$$w = \Phi(z) = \frac{1}{\varphi\left(\dfrac{1}{z - z_0}\right)},$$

mapping G_∞ onto the domain $|w| > \rho > 0$, has the desired properties (3.22). In fact

$$\Phi(\infty) = \frac{1}{\varphi(0)} = \infty,$$

while

$$\lim_{z \to \infty} \frac{\Phi(z)}{z} = \lim_{z_1 \to 0} \frac{1}{z_0 z_1 + 1} \frac{z_1}{w_1} = 1.$$

COROLLARY. *The function $\Phi(z)$ figuring in Theorem* 3.14 *has a Laurent expansion of the form*

$$\Phi(z) = z + \alpha_0 + \frac{\alpha_{-1}}{z} + \cdots$$

---

[18] G_∞ is a simply connected domain in the extended plane, whose boundary Γ_∞ is contained in F.

at infinity. Moreover, given any integer $n > 0$, the function $[\Phi(z)]^n$ has a Laurent expansion of the form

$$[\Phi(z)]^n = z^n + \alpha_{n-1}^{(n)} z^{n-1} + \cdots + \alpha_0^{(n)} + \frac{\alpha_{-1}^{(n)}}{z} + \cdots \qquad (3.23)$$

at infinity.

The polynomials

$$\Phi_n(z) = z^n + \alpha_{n-1}^{(n)} z^{n-1} + \cdots + \alpha_0^{(n)}$$

consisting of the terms with nonnegative powers of z in the expansion (3.23) are called the *Faber polynomials* generated by the continuum F, or simply the Faber polynomials for F. It is obvious that

$$\Phi_0(z) = 1, \qquad \Phi_1(z) = z + \alpha_0, \qquad \Phi_2(z) = z^2 + 2\alpha_0 z + a_0^2 + 2\alpha_{-1}, \ldots$$

Example 1. If F is the closed disk $|z - z_0| \leqslant \rho$, then

$$w = \Phi(z) = z - z_0,$$

and hence

$$\Phi_n(z) = (z - z_0)^n.$$

Example 2. Let F be the closed interior of the lemniscate

$$|z^k + A_{k-1} z^{k-1} + \cdots + A_0| = \rho^k,$$

with k foci.[19] Then it is easy to see that

$$\Phi(z) = z\left(1 + \frac{A_{k-1}}{z} + \cdots + \frac{A_0}{z^k}\right)^{1/k},$$

where we choose the branch of the multiple-valued function

$$\left(1 + \frac{A_{k-1}}{z} + \cdots + \frac{A_0}{z^k}\right)^{1/k}$$

which equals 1 for $z = \infty$. Therefore

$$[\Phi(z)]^{mk} = (z^k + A_{k-1} z^{k-1} + \cdots + A_0)^m \qquad (m = 0, 1, 2, \ldots),$$

i.e.,

$$\Phi_{mk}(z) = (z^k + A_{k-1} z^{k-1} + \cdots + A_0)^m.$$

In particular, for the lemniscate

$$|z^2 - 1| = 1$$

with two foci, we have

$$\Phi(z) = z\left(1 - \frac{1}{z^2}\right)^{1/2},$$

---

[19] In the present case, the lemniscate can have only one component, as described on p. I.379. If the lemniscate intersects itself, F is the union of several closed domains.

and

$$\Phi_0(z) = 1, \quad \Phi_1(z) = z, \quad \Phi_2(z) = z^2 - 1, \quad \Phi_3(z) = z^3 - \tfrac{3}{2}z,$$
$$\Phi_4(z) = z^4 - 2z^2 + 1, \qquad \Phi_5(z) = z^5 - \tfrac{5}{2}z^3 + \tfrac{15}{8}z, \ldots$$

Example 3. Let F be the segment $-1 \leqslant x \leqslant 1$ of the real axis. Then

$$w = \Phi(z) = \tfrac{1}{2}(z + \sqrt{z^2 - 1})$$

maps G_∞ onto the domain $|w| > \frac{1}{2}$ (cf. Sec. I.51).[20] Since the Laurent expansion at infinity of the function

$$\frac{1}{4\Phi(z)} = \tfrac{1}{2}(z - \sqrt{z^2 - 1})$$

contains no nonnegative powers of z, the function

$$[\Phi(z)]^n + \frac{1}{[4\Phi(z)]^n} = [\tfrac{1}{2}(z + \sqrt{z^2 - 1})]^n + [\tfrac{1}{2}(z - \sqrt{z^2 - 1})]^n \tag{3.24}$$

has the same Laurent expansion at infinity as $[\Phi(z)]^n$ itself if $n = 1, 2, \ldots$ But (3.24) is a polynomial of degree n, and hence

$$\Phi_0(z) = 1, \quad \Phi_n(z) = \frac{1}{2^n}[(z + \sqrt{z^2 - 1})^n + (z - \sqrt{z^2 - 1})^n] \qquad (n = 1, 2, \ldots).$$

Setting $z = \cos t$, we obtain

$$\Phi_n(\cos t) = \frac{1}{2^n}[(\cos t + i \sin t)^n + (\cos t - i \sin t)^n] = \frac{1}{2^{n-1}} \cos nt,$$

or equivalently,

$$\Phi_0(z) = 1, \quad \Phi_n(z) = \frac{1}{2^{n-1}} \cos (n \operatorname{Arc} \cos z) = T_n(z) \qquad (n = 1, 2, \ldots).$$

Thus, if F is the interval $[-1, 1]$ of the real axis, the Faber polynomials turn out to be the (suitably normalized) classical Chebyshev polynomials which deviate least from zero on $[-1, 1]$ (see Problems 3.15–3.19).

THEOREM 3.15. *With the same notation as in Theorem* 3.14, *let* $\Psi(w)$ *be the inverse of* $\Phi(z)$, *and suppose* F *is contained in the disk* $|z| < R$. *Then, if* $\Phi_n(z)$, $n = 0, 1, 2, \ldots$ *are the Faber polynomials for* F,

$$\frac{\Psi'(w)}{\Psi(w) - z} = \sum_{n=0}^{\infty} \frac{\Phi_n(z)}{w^{n+1}}, \tag{3.25}$$

where the convergence is uniform for all $|w| \geqslant R$.

---

[20] Here we choose the branch of $\sqrt{z^2 - 1}$ such that

$$\lim_{z \to \infty} \frac{\sqrt{z^2 - 1}}{z} = 1.$$

Proof. Clearly,

$$\frac{1}{2\pi i}\int_{|\zeta|=R}\frac{[\Phi(\zeta)]^n - \Phi_n(\zeta)}{\zeta - z}\,d\zeta = 0 \qquad (n = 0, 1, 2, \ldots) \tag{3.26}$$

if $|z| < R$. In fact, the integrand is analytic for $|\zeta| \geqslant R$ and has a zero of at least the second order at ∞, since $[\Phi(\zeta)]^n - \Phi_n(\zeta)$ has a zero of at least the first order at ∞. Therefore the residue of the integrand at ∞ vanishes, which implies (3.26). It follows from (3.26) and Cauchy's integral formula that

$$\Phi_n(z) = \frac{1}{2\pi i}\int_{|\zeta|=R}\frac{[\Phi(\zeta)]^n}{\zeta - z}\,d\zeta, \tag{3.27}$$

where the circle $|\zeta| = R$ can be replaced by any closed rectifiable Jordan curve γ such that $\gamma \subset G_\infty$, $z \in I(\gamma)$. In particular, we can choose γ to be a *circular image* C_R, i.e., the inverse image under $w = \Phi(z)$ of a circle $|w| = R > \rho$. Then (3.27) becomes

$$\Phi_n(z) = \frac{1}{2\pi i}\int_{C_R}\frac{[\Phi(\zeta)]^n}{\zeta - z}\,d\zeta = \frac{1}{2\pi i}\int_{|w|=R}\frac{\Psi'(w)w^n}{\Psi(w) - z}\,dw, \tag{3.28}$$

where $\Psi(w) = \Phi^{-1}(w)$ and z is any point of $I(C_R)$, in particular any point of F. Therefore $\Phi_n(z)$ is the coefficient of w^{-n-1} $(n = 0, 1, 2, \ldots)$ in the Laurent expansion of the function

$$\chi(w; z) = \frac{\Psi'(w)}{\Psi(w) - z}$$

at infinity. But since $z = \Psi(w)$ has a simple pole at ∞ and since

$$\lim_{w\to\infty}\frac{\Psi(w)}{w} = \lim_{z\to\infty}\frac{z}{\Phi(z)} = 1,$$

we have

$$\Psi(w) = w + \beta_0 + \frac{\beta_{-1}}{w} + \cdots, \quad \Psi'(w) = 1 - \frac{\beta_{-1}}{w^2} + \cdots,$$

and hence $\chi(w; z)$, regarded as a function of w, has a simple zero at infinity. Therefore the expansion (3.25) is uniformly convergent for all $|w| \geqslant R$, as asserted.

Remark. The function $\chi(w; z)$ is called the *generating function* of the Faber polynomials $\Phi_n(z)$, $n = 0, 1, 2, \ldots$

LEMMA. *With the same notation as in Theorem* 3.15 *and its proof,*

$$\Phi_n(z) = [\Phi(z)]^n + \frac{l_r}{2\pi\delta_{r,R'}}\,r^n\vartheta_n(z; r, R'), \qquad z \in \overline{E(C_{R'})}, \quad z \neq \infty,$$

where r and R' are any numbers such that $\rho < r < R'$, l_r is the length of C_r, $\delta_{r,R'}$ is the distance between C_r and $C_{R'}$, and $|\vartheta(z; r, R')| < 1$.[21] *In particular,*

$$\tfrac{1}{2}|\Phi(z)|^n < |\Phi_n(z)| < \tfrac{3}{2}|\Phi(z)|^n, \qquad z \in \overline{E(C_{R'})}, \quad z \neq \infty \tag{3.30}$$

for all sufficiently large $n > N(r, R')$, and

$$\lim_{n\to\infty} \sqrt[n]{|\Phi_n(z)|} = |\Phi(z)|. \tag{3.31}$$

Proof. According to (3.28), if z is any finite point of $E(C_{R'})$, then

$$\Phi_n(z) = \frac{1}{2\pi i}\int_{C_R} \frac{[\Phi(\zeta)]^n}{\zeta - z}\,d\zeta = [\Phi(z)]^n + \frac{1}{2\pi i}\int_{C_r} \frac{[\Phi(\zeta)]^n}{\zeta - z}\,d\zeta \tag{3.32}$$

for sufficiently large R. But

$$\left|\frac{1}{2\pi i}\int_{C_r} \frac{[\Phi(\zeta)]^n}{\zeta - z}\,d\zeta\right| < \frac{l_r}{2\pi\delta_{r,R'}}\,r^n,$$

which together with (3.32) implies (3.29). To derive (3.30) from (3.29), we observe that $|\Phi(z)| \geqslant R'$ on $\overline{E(C_{R'})}$, while

$$\frac{l_r}{2\pi\delta_{r,R'}}\frac{r^n}{R'^n} < \frac{1}{2}$$

for all sufficiently large $n > N(r, R')$. Finally we note that (3.31) is an obvious consequence of (3.30).

We now consider expansions in series of Faber polynomials, i.e., expansions of the form

$$\sum_{n=0}^{\infty} a_n\Phi_n(z). \tag{3.33}$$

Our first result is the analogue of the Cauchy-Hadamard theorem for power series (Theorem I.16.1):

THEOREM 3.16. *If*

$$\overline{\lim_{n\to\infty}} \sqrt[n]{|a_n|} = \frac{1}{R_0} > \frac{1}{\rho}, \tag{3.34}$$

then the series (3.33) *converges absolutely and uniformly inside $I(C_{R_0})$, and diverges on $E(C_{R_0})$. In particular,* (3.33) *represents an analytic function on $I(C_{R_0})$.*

Proof. Choosing $\rho < R' < R_0$ in (3.30), we find that

$$|\Phi_n(z)| < \tfrac{3}{2}|\Phi(z)|^n = \tfrac{3}{2}R'^n \qquad (z \in C_{R'})$$

---

[21] $\overline{E(C_{R'})}$ means the closed exterior of the circular image $C_{R'}$.

for all sufficiently large $n > N(R')$ [r can be set equal to $\frac{1}{2}(\rho + R')$]. On the other hand, given any ε such that $0 < \varepsilon < R_0 - R'$, we have

$$\sqrt[n]{|a_n|} < \frac{1}{R_0 - \frac{\varepsilon}{2}} < \frac{1}{R' + \frac{\varepsilon}{2}}$$

for all sufficiently large $n > N'(\varepsilon)$. Therefore, if $n > \nu = \max(N, N')$, we have

$$|a_n|\,|\Phi_n(z)| < \frac{3}{2}\left(\frac{R'}{R' + \frac{\varepsilon}{2}}\right)^n$$

for all $z \in C_{R'}$, and hence for all $z \in \overline{I(C_{R'})}$. But then, since $R' < R_0$ is arbitrary, the series (3.33) is absolutely and uniformly convergent inside $I(C_{R_0})$, which incidentally implies the last assertion of the theorem. However, if z_0 belongs to $E(C_{R_0})$, then $|\Phi(z)| > R_0$, and hence, because of (3.31) and (3.34),

$$\overline{\lim_{n\to\infty}}\ \sqrt[n]{|a_n\Phi_n(z)|} = \frac{|\Phi(z)|}{R_0} > 1.$$

i.e., the series (3.33) diverges.

Our next theorem is the analogue of Cauchy's theorem on the expansion of an analytic function in power series (Theorem I.16.7):

THEOREM 3.17. *Every function $f(z)$ analytic on $I(C_{R_0})$, where $R_0 > \rho$, can be represented on $I(C_{R_0})$ as the sum of a series of the form* (3.33).

Proof. Given any $z \in I(C_{R_0})$, there is a number R such that $\rho < R < R_0$ and $z \in I(C_R)$. Then we have

$$f(z) = \frac{1}{2\pi i}\int_{C_R} \frac{f(\zeta)}{\zeta - z}\,d\zeta = \frac{1}{2\pi i}\int_{|w|=R} \frac{f[\Psi(w)]\Psi'(w)}{\Psi(w) - z}\,dw. \quad (3.35)$$

Replacing

$$\frac{\Psi'(w)}{\Psi(w) - z}$$

by its expansion (3.25) which is uniformly convergent on the circle $|w| = R$, we obtain the series (3.33) with coefficients

$$a_n = \frac{1}{2\pi i}\int_{|w|=R} \frac{f[\Psi(w)]}{w^{n+1}}\,dw = \frac{1}{2\pi i}\int_{C_R} \frac{f(z)\Phi'(z)}{[\Phi(z)]^{n+1}}\,dz. \quad (3.36)$$

COROLLARY. *If $f(z)$ is analytic on $I(C_{R_0})$, then the coefficients a_n of its series representation* (3.33) *satisfy the inequalities*

$$|a_n| \leqslant \frac{M(R)}{R^n} \qquad (n = 0, 1, 2, \ldots), \quad (3.37)$$

where

$$M(R) = \max_{z \in C_R} |f(z)| \qquad (R < R_0).$$

Proof. This result, which is the analogue of the familiar Cauchy inequalities (see p. I.358) is an immediate consequence of (3.36).

Remark. It follows from (3.37) that

$$\overline{\lim_{n \to \infty}} \sqrt[n]{|a_n|} \leqslant \frac{1}{R} \qquad (R < R_0),$$

and hence that

$$\overline{\lim_{n \to \infty}} \sqrt[n]{|a_n|} \leqslant \frac{1}{R_0},$$

since $R < R_0$ is arbitrary. Therefore, according to Theorem 3.16, the series (3.33), with coefficients (3.36), converges absolutely and uniformly inside $I(C_{R_0})$.

Finally, we prove that expansions in Faber polynomials have the uniqueness property on $I(C_{R_0})$:[22]

THEOREM 3.18. *If the sums of two series of the form* (3.33) *coincide on* $I(C_{R_0})$, *where* $R_0 > \rho$, *then identical Faber polynomials have identical coefficients, i.e., there is a unique series in the Faber polynomials which has a given sum on* $I(C_{R_0})$.

Proof. Forming the difference between the two functions whose sums coincide on $I(C_{R_0})$, we obtain a series

$$\sum_{n=0}^{\infty} a_n \Phi_n(z) = 0, \qquad z \in I(C_{R_0}), \tag{3.38}$$

where

$$\overline{\lim_{n \to \infty}} \sqrt[n]{|a_n|} \leqslant \frac{1}{R_0}. \tag{3.39}$$

Let R' be such that $\rho < R' < R_0$, and choose $r = \frac{1}{2}(\rho + R')$. Then, according to (3.29),

$$|[\Phi(z)]^n - \Phi_n(z)| < \frac{l_r}{2\pi\delta_{r,R'}} r^n$$

for all $z \in \overline{E(C_{R'})}$, including $z = \infty$, provided that $n > N'(R')$. It follows that the series

$$\sum_{n=0}^{\infty} a_n\{[\Phi(z)]^n - \Phi_n(z)\} \tag{3.40}$$

is uniformly convergent on $\overline{E(C_{R'})}$, and hence represents an analytic function on $\overline{E(C_{R'})}$, which vanishes at ∞, since all the terms of (3.40)

[22] See p. I.350, esp. Theorem I.16.5.

vanish at ∞. Carrying out the transformation $w = \Phi(z)$, we find that the function

$$\varphi(w) = \sum_{n=0}^{\infty} a_n\{w^n - \Phi_n[\Psi(w)]\} \tag{3.41}$$

is analytic for $|w| > R'$ (and hence for $|w| > \rho$, since $R' > \rho$ is arbitrary), and vanishes at ∞. According to (3.38), for $\rho < |w| < R_0$ we have

$$\sum_{n=0}^{\infty} a_n\Phi_n[\Psi(w)] = 0,$$

i.e.,

$$\varphi(w) = \sum_{n=0}^{\infty} a_n w^n. \tag{3.42}$$

But this power series converges for $|w| < R_0$, because of (3.39). Therefore the function $\varphi(w)$ defined by the series (3.41) for $|w| > \rho$ and by the series (3.42) for $|w| < R_0$ is analytic in the whole plane, and moreover vanishes at ∞. It follows from Liouville's theorem that $\varphi(w) \equiv 0$, and hence from (3.42) that

$$a_n = 0 \qquad (n = 0, 1, 2, \ldots),$$

as required.

Example 1. Let the continuum F be the closed disk $|z - z_0| \leqslant \rho$. Then the Faber polynomials are

$$\Phi_n(z) = (z - z_0)^n \qquad (n = 0, 1, 2, \ldots),$$

the circular images C_r are the circles $|z - z_0| = r$, and the series (3.33) reduces to the Taylor series

$$\sum_{n=0}^{\infty} a_n(z - z_0)^n.$$

Example 2. Let F be the segment $-1 \leqslant x \leqslant 1$ of the real axis. Then, as in Example 3, p. 106,

$$\Phi_n(z) = T_n(z) = \frac{1}{2^{n-1}} \cos (n \operatorname{Arc} \cos z) \qquad (n = 0, 1, 2, \ldots),$$

where the $T_n(z)$ are the Chebyshev polynomials. The circular images are now the confocal ellipses

$$\frac{x^2}{\left(r + \frac{1}{4r}\right)^2} + \frac{y^2}{\left(r - \frac{1}{4r}\right)^2} = 1 \tag{3.43}$$

with foci ± 1 (cf. p. I.198). According to Theorems 3.17 and 3.18, every

function $f(z)$ analytic on $I(C_{R_0})$, where $R_0 > \rho = \frac{1}{2}$, can be represented by a unique series of Chebyshev polynomials

$$f(z) = \sum_{n=0}^{\infty} a_n T_n(z),$$

which converges uniformly inside $I(C_{R_0})$.

15. Bernstein's Theorem

The Faber polynomials can be used to study polynomial approximation of the functions defined on a continuum F. We now give some results first obtained by Bernstein in his famous dissertation (1912), for the case where F is a segment of the real axis.

LEMMA (*Bernstein's lemma*). *Let F be a continuum containing more than one point, and let G_∞ be the component of F^c containing the point at infinity. Suppose the function*

$$z = \Psi(w) = w + \beta_0 + \frac{\beta_{-1}}{w} + \cdots$$

maps the domain $|w| < \rho$ conformally onto G_∞, and let C_R be the image of the circle $|w| = R > \rho$ under the mapping $z = \Psi(w)$. Suppose $\Pi_n(z)$ is a polynomial of degree at most n which satisfies the inequality

$$|\Pi_n(z)| \leqslant M$$

for all $z \in F$. Then $\Pi_n(z)$ satisfies the inequality

$$|\Pi_n(z)| \leqslant M\left(\frac{R}{\rho}\right)^n$$

for all $z \in C_R$.[23]

Proof. If $\Phi(z) = \Psi^{-1}(w)$, then the function

$$\varphi(z) = \frac{\Pi_n(z)}{[\Phi(z)]^n}$$

is analytic on the domain G_∞, and moreover

$$\varphi(\infty) = \lim_{z\to\infty} \frac{\Pi_n(z)}{[\Phi(z)]^n} = \lim_{z\to\infty} \left\{\left[\frac{z}{\Phi(z)}\right]^n \frac{\Pi_n(z)}{z^n}\right\} = A_n,$$

where A_n is the coefficient of z^n in the polynomial $\Pi_n(z)$. Applying the maximum modulus principle to $\varphi(z)$ on the closed domain $\overline{E(C_r)}$, where $r > \rho$, we find that

$$|\varphi(z)| \leqslant \max_{\zeta\in C_r} |\varphi(\zeta)| = \frac{\max_{\zeta\in C_r} |\Pi_n(\zeta)|}{r^n}$$

[23] And hence for all $z \in \overline{I(C_R)}$, by the maximum modulus principle.

for all $z \in \overline{E(C_r)}$.[24] In particular, choosing $z \in C_R$ $(R > r)$, so that $|\Phi(z)| = R$, we have

$$|\Pi_n(z)| \leqslant \max_{\zeta \in C_r} |\Pi_n(\zeta)| \left(\frac{R}{r}\right)^n < \max_{\zeta \in C_r} |\Pi_n(\zeta)| \left(\frac{R}{\rho}\right)^n. \tag{3.44}$$

We now take the limit of (3.44) as $r \to \rho+$, obtaining

$$|\Pi_n(z)| \leqslant \lim_{r \to \rho+} \max_{\zeta \in C_r} |\Pi_n(\zeta)| \left(\frac{R}{\rho}\right)^n = \mu \left(\frac{R}{\rho}\right)^n,$$

where the existence of the limit

$$\mu = \lim_{r \to \rho+} \max_{\zeta \in C_r} |\Pi_n(\zeta)|$$

follows from the fact that

$$\max_{\zeta \in C_r} |\Pi_n(\zeta)|$$

is nonincreasing as $r \to \rho+$. Clearly, if $\zeta_r \in C_r$ is such that

$$\max_{\zeta \in C_r} |\Pi_n(\zeta)| = |\Pi_n(\zeta_r)|,$$

then

$$\mu = \lim_{r \to \rho+} |\Pi_n(\zeta_r)|,$$

or equivalently,

$$\mu = \lim_{k \to \infty} |\Pi_n(\zeta_{r_k})|,$$

where $\{r_k\}$ is any sequence converging to ρ $(r_k > \rho)$. Applying Theorem 2.5, we see that all the limit points of $\{\zeta_{r_k}\}$ belong to the boundary of G_∞ and hence to F. Therefore, going over to a subsequence if necessary, we can write

$$\lim_{k \to \infty} \zeta_{r_k} = z_0 \in F.$$

But then

$$\mu = \lim_{k \to \infty} |\Pi_n(\zeta_{r_k})| = |\Pi_n(z_0)| \leqslant M,$$

i.e.,

$$|\Pi_n(z)| \leqslant \mu \left(\frac{R}{\rho}\right)^n \leqslant M \left(\frac{R}{\rho}\right)^n,$$

as asserted.

---

[24] To map $E(C_r)$ onto a bounded domain, so that the maximum modulus principle can be applied in its usual form, we use the transformation

$$Z = \frac{1}{z - z_0},$$

where z_0 is any point of the continuum F.

THEOREM 3.19 (*Bernstein's theorem*). *Let F and ρ be the same as in Bernstein's lemma, and let $f(z)$ be a function defined on F. Then, given any $\varepsilon > 0$ and any integer $n \geqslant 0$, there exists a polynomial $\Pi_n(z)$ of degree at most n such that*

$$|f(z) - \Pi_n(z)| < c(\varepsilon)(q + \varepsilon)^n \qquad (0 < q < 1) \tag{3.45}$$

for all $z \in F$ if and only if $f(z)$ is analytic on the domain $I(C_{R_0})$, where $R_0 = \rho/q$. In this case, the sequence $\{\Pi_n(z)\}$ converges uniformly to $f(z)$ inside $I(C_{R_0})$.

Proof. First suppose $f(z)$ is analytic on $I(C_{R_0})$, where $R_0 = \rho/q$ and $0 < q < 1$. Then, according to Theorem 3.17, $f(z)$ can be represented by a series of Faber polynomials

$$f(z) = \sum_{k=0}^{\infty} a_k \Phi_k(z)$$

which converges uniformly inside $I(C_{R_0})$. The partial sum

$$\Pi_n(z) = \sum_{k=0}^{n} a_k \Phi_k(z)$$

of this series is a polynomial of degree at most n, since each $\Phi_k(z)$ is of degree k. We now show that the polynomials $\Pi_n(z)$, $n = 0, 1, 2, \ldots$ satisfy the inequality (3.45). To this end, let ε be any number such that $0 < \varepsilon < 1 - q$, and then choose R' and R such that $\rho < R' < R < R_0$ and

$$\frac{R'}{R} = q + \varepsilon \quad \left(> \frac{\rho}{R_0}\right)$$

(obviously, this is always possible). But, according to (3.37),

$$|a_k| \leqslant \frac{M(R)}{R^k},$$

where

$$M(R) = \max_{z \in C_R} |f(z)|,$$

and moreover, according to (3.30),

$$|\Phi_k(z)| < \tfrac{3}{2}|\Phi(z)|^k = \tfrac{3}{2}R'^k$$

for all $z \in C_{R'}$ and all sufficiently large $n > N(\varepsilon)$ [r can be set equal to $\frac{1}{2}(\rho + R')$]. Therefore

$$|f(z) - \Pi_n(z)| = \left| \sum_{k=n+1}^{\infty} a_k \Phi_k(z) \right| < \tfrac{3}{2}M(R) \sum_{k=n+1}^{\infty} \left(\frac{R'}{R}\right)^k$$

$$= \tfrac{3}{2}M(R) \frac{(R'/R)^{n+1}}{1 - (R'/R)} = \tfrac{3}{2}M(R) \frac{q + \varepsilon}{1 - (q + \varepsilon)} (q + \varepsilon)^n$$

for all $z \in C_{R'}$ and $n > N(\varepsilon)$. By the maximum modulus principle, this inequality also holds at every point of the continuum F. Replacing the number

$$\tfrac{3}{2}M(R)\frac{q+\varepsilon}{1-(q+\varepsilon)} \tag{3.46}$$

by a larger number if necessary, we can make the inequality hold for $n \leqslant N(\varepsilon)$ as well. In other words, denoting (3.46) or this larger number by $c(\varepsilon)$, we have

$$|f(z) - \Pi_n(z)| < c(\varepsilon)(q+\varepsilon)^n$$

for all $z \in F$ and $n \geqslant 0$, which agrees with (3.45). This proves the sufficiency part of Bernstein's theorem.

To prove the necessity, let $f(z)$ be a function defined on F, and suppose that given any $\varepsilon > 0$ and any $n \geqslant 0$, there exists a polynomial $\Pi_n(z)$ of degree at most n satisfying (3.45) for all $z \in F$. Then obviously the sequence $\{\Pi_n(z)\}$ converges uniformly to $f(z)$ on F. But $\{\Pi_n(z)\}$ also converges uniformly inside $I(C_{R_0})$, where $R_0 = \rho/q$. To see this, we first compare (3.45) with the same inequality where n is replaced by $n-1$, obtaining

$$|\Pi_n(z) - \Pi_{n-1}(z)| < c(\varepsilon)(q+\varepsilon)^n\left(1+\frac{1}{q+\varepsilon}\right) = c'(\varepsilon)(q+\varepsilon)^n$$

for all $z \in F$ [$\Pi_{-1}(z) \equiv 0$]. Then, choosing

$$\varepsilon < \frac{1-q}{2}, \qquad R = \frac{\rho}{q+2\varepsilon} \quad (>\rho),$$

we apply Bernstein's lemma to the polynomial $\Pi_n(z) - \Pi_{n-1}(z)$, of degree at most n, deducing that

$$|\Pi_n(z) - \Pi_{n-1}(z)| < c'(\varepsilon)(q+\varepsilon)^n\left(\frac{R}{\rho}\right)^n = c'(\varepsilon)\left(\frac{q+\varepsilon}{q+2\varepsilon}\right)^n$$

for all $z \in C_R$. It follows that the series

$$\Pi_0(z) + [\Pi_1(z) - \Pi_0(z)] + \cdots + [\Pi_n(z) - \Pi_{n-1}(z)] + \cdots$$

is uniformly convergent on $\overline{I(C_R)}$, and hence the same is true of the sequence $\{\Pi_n(z)\}$. But

$$\lim_{\varepsilon \to 0} R = R_0,$$

and hence the sequence $\{\Pi_n(z)\}$ is uniformly convergent inside $I(C_{R_0})$. Therefore the function

$$\lim_{n \to \infty} \Pi_n(z),$$

which coincides with $f(z)$ on F, is analytic on $I(C_{R_0})$, and the proof is complete.

Example 1. If F is the closed disk $|z - z_0| \leqslant \rho$, then the existence of a polynomial $\Pi_n(z)$ of degree at most n satisfying the inequality

$$|f(z) - \Pi_n(z)| < c(\varepsilon)(q + \varepsilon)^n \qquad (0 < q < 1,\ z \in F) \tag{3.47}$$

(given any $\varepsilon > 0$ and any integer $n \geqslant 0$) is necessary and sufficient for $f(z)$ to be analytic on the open disk $|z - z_0| < \rho/q$.

Example 2. If F is the segment $-1 \leqslant x \leqslant 1$ of the real axis, the condition (3.47) is necessary and sufficient for $f(z)$ to be analytic on the interior of the ellipse

$$\frac{x^2}{\left(\frac{1}{2q} + \frac{q}{2}\right)^2} + \frac{y^2}{\left(\frac{1}{2q} - \frac{q}{2}\right)^2} = 1$$

[cf. (3.43)]. This is an ellipse with foci ± 1, the sum of whose semiaxes is $1/q$.

16. Approximation in the Mean

A function $F(z)$ defined on a domain G is said to be *square-integrable* on G if[25]

$$\iint_G |F(z)|^2\, dx\, dy < \infty. \tag{3.48}$$

If, given any $\varepsilon > 0$, there exists a polynomial $\Pi(z)$ such that

$$\iint_G |F(z) - \Pi(z)|^2\, dx\, dy < \varepsilon,$$

we say that $F(z)$ can be *approximated in the mean on* G (more exactly, *with respect to the area of* G) by polynomials. Obviously, in this case there exists a sequence of polynomials $\{\Pi_n(z)\}$ such that

$$\lim_{n\to\infty} \iint_G |F(z) - \Pi_n(z)|^2\, dx\, dy = 0,$$

and then we say that $\{\Pi_n(z)\}$ *converges to* $F(z)$ *in the mean* (*on* G).

---

[25] The integral in (3.48) is interpreted as an improper integral in the following sense: Let $\{D_n\}$ be an increasing sequence of bounded domains (each with a well-defined area) approximating G (cf. p. 81), i.e., $\bar{D}_n \subset G$, $\bar{D}_n \subset D_{n+1}$ $(n = 1, 2, \ldots)$ and $E \subset D_n$ for any compact set $E \subset G$ and all $n > N(E, \varepsilon)$. (Such a sequence of domains was constructed in the proof of Theorem 1.10.) Then

$$\iint_G |F(z)|^2\, dx\, dy \equiv \lim_{n\to\infty} \iint_{\bar{D}_n} |F(z)|^2\, dx\, dy,$$

and similarly for other integrals over G (see Problem 3.21). Here, of course, it is assumed that the integrals over $\bar{D}_n$ all exist (this is the case, for example, if $F(z)$ is continuous or analytic on G).

THEOREM 3.20. *If $F(z)$ is analytic and square-integrable on a Carathéodory domain G, then $F(z)$ can be approximated in the mean on G by polynomials.*

Proof. Since $F(z)$ is square-integrable on G, we have

$$\iint_{|w|<1} |F^*(w)[f^{-1}(w)]'|^2 \, du \, dv = \iint_G |F(z)|^2 \, dx \, dy = C < \infty,$$

where $f^{-1}(w)$ maps the disk $|w| < 1$ conformally onto G and $F^*(w) = F[f^{-1}(w)]$. If

$$\varphi(w) = F^*(w)[f^{-1}(w)]' = \sum_{k=0}^{\infty} a_k w^k \qquad (|w| < 1),$$

then

$$\begin{aligned}\iint_{|w|\leqslant r<1} |\varphi(w)|^2 \, du \, dv &= \int_0^{2\pi} \int_0^r \left| \sum_{k=0}^{\infty} a_k w^k \right|^2 r \, dr \, d\theta \\ &= \int_0^r dr \left[\int_0^{2\pi} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} a_k \bar{a}_l r^{k+l+1} e^{i(k-l)\theta} \, d\theta \right] \\ &= 2\pi \int_0^r \sum_{k=0}^{\infty} |a_k|^2 r^{2k+1} \, dr = \pi \sum_{k=0}^{\infty} |a_k|^2 \frac{r^{2k+2}}{k+1} < C.\end{aligned}$$

Therefore

$$\pi \sum_{k=0}^{N} |a_k|^2 \frac{r^{2k+2}}{k+1} < C$$

for every N and $r < 1$, which implies

$$\pi \sum_{k=0}^{N} \frac{|a_k|^2}{k+1} \leqslant C,$$

and hence

$$\pi \sum_{k=0}^{\infty} \frac{|a_k|^2}{k+1} \leqslant C,$$

i.e., the series

$$\sum_{k=0}^{\infty} \frac{|a_k|^2}{k+1} \tag{3.49}$$

converges.

Next consider the integral

$$\begin{aligned}\iint_{|w|\leqslant r<1} \left| \varphi(w) - \sum_{k=0}^{N} a_k w^k \right|^2 du \, dv &= \iint_{|w|\leqslant r<1} \left| \sum_{k=N+1}^{\infty} a_k w^k \right|^2 du \, dv \\ &= \pi \sum_{k=N+1}^{\infty} |a_k|^2 \frac{r^{2k+2}}{k+1} \\ &< \pi \sum_{k=N+1}^{\infty} \frac{|a_k|^2}{k+1}. \end{aligned} \tag{3.50}$$

Taking the limit of (3.50) as $r \to 1-$, we obtain

$$\iint_{|w|<1} \left|\varphi(w) - \sum_{k=0}^{N} a_k w^k\right|^2 du\, dv \leqslant \pi \sum_{k=N+1}^{\infty} \frac{|a_k|^2}{k+1}.$$

But then

$$\lim_{n\to\infty} \iint_{|w|<1} \left|\varphi(w) - \sum_{k=0}^{N} a_k w^k\right|^2 du\, dv = 0,$$

since (3.49) converges. Using the transformation $w = f(z)$ to return to the z-plane, we find that

$$\iint_{|w|<1} \left|\varphi(w) - \sum_{k=0}^{N} a_k w^k\right|^2 du\, dv = \iint_G \left|\frac{F(z)}{f'(z)} - \sum_{k=0}^{N} a_k [f(z)]^k\right|^2 |f'(z)|^2\, dx\, dy$$
$$= \iint_G \left|F(z) - \sum_{k=0}^{N} a_k [f(z)]^k f'(z)\right|^2 dx\, dy,$$

and hence

$$\lim_{N\to\infty} \iint_G \left|F(z) - \sum_{k=0}^{N} a_k [f(z)]^k f'(z)\right|^2 dx\, dy = 0.$$

The next step is to use Runge's theorem to show that the functions

$$\sum_{k=0}^{N} a_k [f(z)]^k f'(z),$$

and hence $F(z)$ itself, can be approximated in the mean by polynomials. To this end, given any $\varepsilon > 0$, let $N_0 = N_0(\varepsilon)$ be such that

$$\iint_G \left|F(z) - \sum_{k=0}^{N_0} a_k [f(z)]^k f'(z)\right|^2 dx\, dy < \frac{\varepsilon}{16}.$$

Using Theorem 2.3, we construct a decreasing sequence of bounded simply connected domains $\{G_n\}$ converging to G as its kernel (relative to any point $z_0 \in G$), and then we let $w = f_n(z)$, $n = 1, 2, \ldots$ and $w = f(z)$ have the same meaning as in Theorem 2.4. It is clear that[26]

$$\iint_G \left|\sum_{k=0}^{N_0} a_k [f(z)]^k f'(z) - \sum_{k=0}^{N_0} a_k [f_n(z)]^k f_n'(z)\right|^2 dx\, dy$$
$$\leqslant (N_0 + 1) \sum_{k=0}^{N_0} |a_k|^2 \iint_G |[f(z)]^k f'(z) - [f_n(z)]^k f_n'(z)|^2\, dx\, dy. \tag{3.51}$$

[26] Use the inequality

$$\left|\sum_{k=0}^{m} a_k\right|^2 \leqslant (m+1) \sum_{k=0}^{m} |a_k|^2,$$

obtained by setting $b_0 = \cdots = b_m = 1$ in the Cauchy-Schwarz inequality

$$\left|\sum_{k=0}^{m} a_k b_k\right|^2 \leqslant \sum_{k=0}^{m} |a_k|^2 \sum_{k=0}^{m} |b_k|^2.$$

Moreover, according to Theorem 2.4, there is an integer $n_0(\varepsilon) > 0$ such that the right-hand side of (3.51) is less than $\varepsilon/16$ if $n = n_0(\varepsilon)$. As a result, we find that

$$\begin{aligned}\iint_G \left| F(z) - \sum_{k=0}^{N_0} a_k[f_{n_0}(z)]^k f'_{n_0}(z) \right|^2 dx\,dy \\ \leqslant 2 \iint_G \left| F(z) - \sum_{k=0}^{N_0} a_k[f(z)]^k f'(z) \right|^2 dx\,dy \\ + 2 \iint_G \left| \sum_{k=0}^{N_0} a_k[f(z)]^k f'(z) - \sum_{k=0}^{N_0} a_k[f_{n_0}(z)]^k f'_{n_0}(z) \right|^2 dx\,dy < \frac{\varepsilon}{4}.\end{aligned} \tag{3.52}$$

But $\bar{G} \subset G_{n_0}$ for every n_0, by the construction of the sequence $\{G_n\}$. Therefore, applying Runge's theorem (Theorem 3.5) to the function

$$\sum_{k=0}^{N_0} a_k[f_{n_0}(z)]^k f'_{n_0}(z),$$

which is analytic on the bounded simply connected domain G_{n_0}, we can find a polynomial $\Pi(z)$ such that

$$\left| \sum_{k=0}^{N_0} a_k[f_{n_0}(z)]^k f'_{n_0}(z) - \Pi(z) \right|^2 < \frac{\varepsilon}{4S} \qquad (z \in \bar{G}),$$

where S denotes the area of G. It follows that

$$\iint_G \left| \sum_{k=0}^{N_0} a_k[f_{n_0}(z)]^k f'_{n_0}(z) - \Pi(z) \right|^2 dx\,dy < \frac{\varepsilon}{4}. \tag{3.53}$$

Comparing (3.52) and (3.53), we obtain

$$\iint_G |F(z) - \Pi(z)|^2 < \varepsilon.$$

Since ε can be made arbitrarily small, the proof is now complete.

Remark. Theorem 3.20 can be extended to domains more general than Carathéodory domains, resembling Carathéodory domains in a certain precise sense which will not be gone into here.[27]

The class of all functions which are analytic and square-integrable on a bounded domain G will be denoted by $\mathscr{A}(G)$. Thus Theorem 3.20 asserts that if G is a Carathéodory domain, then, given any function $F(z) \in \mathscr{A}(G)$, there exists a sequence of polynomials $\{\Pi_n(z)\}$ converging to $F(z)$ in the mean (on G).

---

[27] The pertinent research, involving the use of delicate function-theoretic techniques, is due to M. V. Keldysh, A. L. Shaginyan and M. M. Dzhrbashyan.

THEOREM 3.21. *If $F(z) \in \mathscr{A}(G)$ and if $\{F_n(z)\}$ is a sequence of functions in $\mathscr{A}(G)$ converging to $F(z)$ in the mean on G,*[28] *then $\{F_n(z)\}$ converges uniformly to $F(z)$ inside G.*

Proof. Given any $z_0 \in G$, let ρ_0 be the distance between z_0 and the boundary of G, and let K be the disk $|z - z_0| < \rho_0$. By hypothesis, given any $\varepsilon > 0$,

$$\iint_G |F(z) - F_n(z)|^2 \, dx \, dy < \varepsilon$$

if $n > N(\varepsilon)$. Therefore, *a fortiori*,

$$\iint_K |F(z) - F_n(z)|^2 \, dx \, dy < \varepsilon$$

if $n > N(\varepsilon)$. Expanding $F(z) - F_n(z)$ in a Taylor series

$$F(z) - F_n(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k \qquad (z \in K),$$

we find that

$$\pi \sum_{k=0}^{\infty} |a_k|^2 \frac{\rho_0^{2k+2}}{k+1} \leqslant \iint_K |F(z) - F_n(z)|^2 \, dx \, dy < \varepsilon,$$

by a calculation like that given at the beginning of the proof of Theorem 3.20. Therefore, in particular,

$$\pi \rho_0^2 |a_0|^2 < \varepsilon.$$

But $a_0 = F(z_0) - F_n(z_0)$, and hence

$$|F(z_0) - F_n(z_0)|^2 < \frac{\varepsilon}{\pi \rho_0^2}$$

if $n > N(\varepsilon)$. This implies the uniform convergence of $\{F_n(z)\}$ to $F(z)$ inside G.

Remark. The converse of Theorem 3.21 is not true. For example, the sequence

$$\{F_n(z)\} = \{\sqrt{n+1}\, z^n\}$$

converges uniformly to zero inside the unit disk. On the other hand,

$$\iint_{|z|<1} |\sqrt{n+1}\, z^n|^2 \, dx \, dy = \pi$$

---

[28] I.e., a sequence of functions such that

$$\lim_{n \to \infty} \iint_G |F(z) - F_n(z)|^2 \, dx \, dy = 0$$

(cf. Problem 3.22).

for all $n = 1, 2, \ldots$, and hence $\{F_n(z)\}$ does not converge to zero in the mean. In fact, according to Theorem 3.21, $\{F_n(z)\}$ cannot converge in the mean to any function, since if $\{F_n(z)\}$ has a limit in the mean, this limit would necessarily vanish on the unit disk.

17. Polynomials Orthogonal on a Domain

A sequence $\{P_n(z)\}$ of polynomials defined on a bounded domain G is said to be *orthogonal on G* (more exactly, *orthogonal with respect to the area of G*) if

$$\iint_G P_m(z)\overline{P_n(z)}\,dx\,dy = 0 \qquad \text{for} \quad m \neq n.$$

If, in addition,

$$\iint_G P_m(z)\overline{P_n(z)}\,dx\,dy = \delta_{mn},$$

where

$$\delta_{mn} = \begin{cases} 1 & \text{for} \quad m = n, \\ 0 & \text{for} \quad m \neq n, \end{cases}$$

the sequence $\{P_n(z)\}$ is said to be *orthonormal on G*.

THEOREM 3.22. *Given a bounded domain G of area S, there is a unique sequence of polynomials $P_0(z), P_1(z), \ldots, P_n(z), \ldots$ orthonormal on G, such that $P_n(z)$ is of degree n and has a positive real leading coefficient.*

Proof. The proof is by induction. If $n = 0$, we obviously have

$$P_0(z) = \frac{1}{\sqrt{S}}.$$

Suppose $P_0(z), P_1(z), \ldots, P_m(z)$ have already been constructed. Then $P_{m+1}(z)$, which can be written in the form

$$P_{m+1}(z) = \mu z^{m+1} + \lambda_m P_m(z) + \cdots + \lambda_0 P_0(z) \qquad (\mu > 0), \quad (3.54)$$

must satisfy the $m + 2$ conditions

$$\iint_G P_{m+1}(z)\overline{P_n(z)}\,dx\,dy = 0 \qquad (n = 0, 1, \ldots, m) \quad (3.55)$$

and

$$\iint_G |P_{m+1}(z)|^2\,dx\,dy = 1. \quad (3.56)$$

These conditions uniquely determine $P_{m+1}(z)$. In fact, substituting (3.54) into (3.55), we find that

$$\lambda_n = -\mu \iint_G z^{m+1}\overline{P_n(z)}\,dx\,dy \equiv -\mu c_n \qquad (n = 0, 1, \ldots, m),$$

so that

$$P_{m+1}(z) = \mu[z^{m+1} - c_m P_m(z) - \cdots - c_0 P_0(z)] \equiv \mu q(z), \tag{3.57}$$

where $q(z) \not\equiv 0$ is a polynomial of degree $m + 1$ whose coefficients are independent of μ. But substitution of (3.57) into (3.56) gives

$$\iint_G |P_{m+1}(z)|^2\,dx\,dy = \mu^2 \iint_G |q(z)|^2\,dx\,dy = 1,$$

and hence

$$\mu = \frac{1}{\sqrt{\iint_G |q(z)|^2\,dx\,dy}},$$

since μ is positive.

Remark. Henceforth, by the *polynomials orthonormal on a domain G*, we shall mean the sequence $\{P_n(z)\}$ constructed in Theorem 3.21.

DEFINITION. *Let $F(z)$ be square-integrable on a domain G, and let $\{P_n(z)\}$ be the polynomials orthonormal on G. Then the numbers*

$$a_n = \iint_G F(z)\overline{P_n(z)}\,dx\,dy \qquad (n = 0, 1, 2, \ldots) \tag{3.58}$$

are called the Fourier coefficients of $F(z)$ with respect to $\{P_n(z)\}$, and the expansion

$$\sum_{n=0}^{\infty} a_n P_n(z) \tag{3.59}$$

is called the Fourier series of $F(z)$ with respect to $\{P_n(z)\}$.

THEOREM 3.23. *Let $F(z)$ be square-integrable on a domain G, with Fourier coefficients* (3.58) *with respect to the polynomials $\{P_n(z)\}$ orthonormal on G. Then the smallest value of the integral*

$$\iint_G \left| F(z) - \sum_{k=0}^{n} c_k P_k(z) \right|^2 dx\,dy$$

is achieved for

$$c_k = a_k \qquad (k = 0, 1, \ldots, n),$$

and is equal to

$$\iint_G |F(z)|^2\,dx\,dy - \sum_{k=0}^{n} |a_k|^2 \geqslant 0. \tag{3.60}$$

Proof. We need only note that

$$\begin{aligned}
&\iint_G \left| F(z) - \sum_{k=0}^{n} c_k P_k(z) \right|^2 \\
&\quad = \iint_G \left[F(z) - \sum_{k=0}^{n} c_k P_k(z) \right] \left[\overline{F(z)} - \sum_{k=0}^{n} \bar{c}_k \overline{P_k(z)} \right] dx\, dy \\
&\quad = \iint_G |F(z)|^2\, dx\, dy - \sum_{k=0}^{n} c_k \iint_G \overline{F(z)} P_k(z)\, dx\, dy \\
&\qquad\qquad - \sum_{k=0}^{n} \bar{c}_k \iint_G F(z) \overline{P_k(z)}\, dx\, dy + \sum_{k=0}^{n} |c_k|^2 \\
&\quad = \iint_G |F(z)|^2\, dx\, dy - \sum_{k=0}^{n} c_k \bar{a}_k - \sum_{k=0}^{n} \bar{c}_k a_k + \sum_{k=0}^{n} c_k \bar{c}_k \\
&\quad = \iint_G |F(z)|^2\, dx\, dy - \sum_{k=0}^{n} |a_k|^2 + \sum_{k=0}^{n} |c_k - a_k|^2 \geqslant 0.
\end{aligned}$$

COROLLARY. *With the same notation as in Theorem* 3.22,

$$\sum_{k=0}^{\infty} |a_k|^2 \leqslant \iint_G |F(z)|^2\, dx\, dy, \tag{3.61}$$

a result known as Bessel's inequality.

Proof. Take the limit as $n \to \infty$ in (3.60).

THEOREM 3.24. *If $\{P_k(z)\}$ are the polynomials orthonormal on a domain G, and if $\{a_k\}$ is a sequence of complex numbers such that*

$$\sum_{k=0}^{\infty} |a_k|^2 \tag{3.62}$$

converges, then

$$\sum_{k=0}^{\infty} a_k P_k(z) \tag{3.63}$$

converges in the mean to a function $F(z) \in \mathscr{A}(G)$, whose Fourier coefficients coincide with the numbers $\{a_k\}$. Moreover

$$\sum_{k=0}^{\infty} |a_k|^2 = \iint_G |F(z)|^2\, dx\, dy, \tag{3.64}$$

a result known as Parseval's theorem.[29]

Proof. Given any $z_0 \in G$, let ρ_0 be the distance between z_0 and the boundary of G, and let K be the disk $|z - z_0| < \rho_0$. Then

$$\iint_K \left| \sum_{k=n+1}^{n+p} a_k P_k(z) \right|^2 dx\, dy \leqslant \iint_G \left| \sum_{k=n+1}^{n+p} a_k P_k(z) \right|^2 dx\, dy = \sum_{k=n+1}^{n+p} |a_k|^2,$$

[29] Cf. pp. I.355 and I.358.

and hence by the argument given in the proof of Theorem 3.21,

$$\left| \sum_{k=n+1}^{n+p} a_k P_k(z_0) \right|^2 \leqslant \frac{1}{\pi \rho_0^2} \sum_{k=n+1}^{n+p} |a_k|^2.$$

Therefore, since (3.62) converges, the series (3.63) converges uniformly inside G to a function $F(z)$ analytic on G. Let $\bar{g}$ be any closed domain contained in G. Then, since (3.63) is uniformly convergent inside G,

$$\iint_{\bar{g}} |F(z)|^2 \, dx \, dy = \lim_{n \to \infty} \iint_{\bar{g}} \left| \sum_{k=0}^{n} a_k P_k(z) \right|^2 dx \, dy.$$

But

$$\begin{aligned} \iint_{\bar{g}} \left| \sum_{k=0}^{n} a_k P_k(z) \right|^2 dx \, dy &\leqslant \iint_{G} \left| \sum_{k=0}^{n} a_k P_k(z) \right|^2 dx \, dy \\ &= \sum_{k=0}^{n} |a_k|^2 \leqslant \sum_{k=0}^{\infty} |a_k|^2, \end{aligned}$$

and hence

$$\iint_{\bar{g}} |F(z)|^2 \, dx \, dy \leqslant \sum_{k=0}^{\infty} |a_k|^2 < \infty.$$

Since $\bar{g} \subset G$ is arbitrary, it follows that

$$\iint_{G} |F(z)|^2 \, dx \, dy \leqslant \sum_{k=0}^{\infty} |a_k|^2 < \infty \tag{3.65}$$

(cf. footnote 25, p. 116), i.e., $F(z) \in \mathscr{A}(G)$. Parseval's theorem (3.64) now follows at once from (3.61) and (3.65).

Next, applying (3.65) to the function

$$\sum_{k=n+1}^{\infty} a_k P_k(z) = F(z) - \sum_{k=0}^{n} a_k P_k(z),$$

we find that

$$\iint_{G} \left| F(z) - \sum_{k=0}^{n} a_k P_k(z) \right|^2 dx \, dy \leqslant \sum_{k=n+1}^{\infty} |a_k|^2. \tag{3.66}$$

Therefore, given any integers n and m such that $n > m \geqslant 0$,

$$\begin{aligned} \left| \iint_{G} F(z) \overline{P_m(z)} \, dx \, dy - a_m \right|^2 &= \left| \iint_{G} \left[F(z) - \sum_{k=0}^{n} a_k P_k(z) \right] \overline{P_m(z)} \, dx \, dy \right|^2 \\ &\leqslant \iint_{G} \left| F(z) - \sum_{k=0}^{n} a_k P_k(z) \right|^2 dx \, dy \iint_{G} |\overline{P_m(z)}|^2 \, dx \, dy \leqslant \sum_{k=n+1}^{\infty} |a_k|^2, \end{aligned} \tag{3.67}$$

where we use the two-dimensional version of Schwarz's inequality. Since the left-hand side of (3.67) is independent of n, while the right-hand side converges to zero as $n \to \infty$, we have

$$a_m = \iint_G F(z)\overline{P_m(z)}\, dx\, dy \qquad (m = 0, 1, 2, \ldots),$$

i.e., (3.63) is the Fourier series of $F(z)$. Then, taking the limit as $n \to \infty$ in (3.66), we find that $F(z)$ is the limit in the mean of its Fourier series, and the proof is complete.

COROLLARY. *A function $F(z) \in \mathscr{A}(G)$ satisfies Parseval's theorem* (3.64) *if and only if it is the limit in the mean of its Fourier series.*

In general, a *given* function $F(z) \in \mathscr{A}(G)$ will not be the limit in the mean of its Fourier series. However, things are particularly simple if G is a Carathéodory domain:

THEOREM 3.25. *If $F(z) \in \mathscr{A}(G)$, where G is a Carathéodory domain, then $F(z)$ is the limit in the mean of its Fourier series.*

Proof. According to Theorem 3.20, there is a sequence of polynomials $\{\Pi_n(z)\}$ converging in the mean to $F(z)$. Without loss of generality, we can assume that $\Pi_n(z)$ is of degree at most n, and hence can be written in the form

$$\Pi_n(z) = c_0 P_0(z) + \cdots + c_n P_n(z).$$

But, according to Theorem 3.23,

$$\iint_G |F(z) - \Pi_n(z)|^2\, dx\, dy \geqslant \iint_G \left| F(z) - \sum_{k=0}^{n} a_k P_k(z) \right|^2 dx\, dy \geqslant 0,$$

where the a_k are the Fourier coefficients of $F(z)$. Therefore, since $\{\Pi_n(z)\}$ converges to $F(z)$ in the mean, so does the sequence

$$\left\{ \sum_{k=0}^{n} a_k P_k(z) \right\}.$$

THEOREM 3.26. *Suppose $w = f(z)$ maps a bounded simply connected domain G conformally onto the unit disk $|w| < 1$ and satisfies the conditions*

$$f(z_0) = 0, \quad f'(z_0) > 0 \qquad (z_0 \in G).$$

Then, for every function $F(z) \in \mathscr{A}(G)$ such that $F(z_0) \neq 0$

$$\iint_G \left| \frac{F(z)}{F(z_0)} \right|^2 dx\, dy \geqslant \iint_G \left| \frac{f'(z)}{f'(z_0)} \right|^2 dx\, dy = \frac{\pi}{|f'(z_0)|^2}, \tag{3.68}$$

where the equality holds if and only if

$$\frac{F(z)}{F(z_0)} = \frac{f'(z)}{f'(z_0)}. \tag{3.69}$$

Proof. Since

$$\iint_G |f'(z)|^2\,dx\,dy = \iint_{|w|<1} du\,dv = \pi,$$

$f'(z)$ obviously belongs to $\mathscr{A}(G)$, and the equality in (3.68) is proved. Clearly, we have

$$\iint_G \left|\frac{F(z)}{F(z_0)}\right|^2 dx\,dy = \iint_{|w|<1} \left|\frac{F[f^{-1}(w)]}{F(z_0)}\right|^2 |[f^{-1}(w)]'|^2\,du\,dv.$$

Writing

$$\varphi(w) = \frac{F[f^{-1}(w)]}{F(z_0)}\,[f^{-1}(w)]' = \sum_{k=0}^{\infty} a_k w^k \qquad (|w| < 1),$$

we find that

$$\iint_G \left|\frac{F(z)}{F(z_0)}\right|^2 dx\,dy \geqslant \pi \sum_{k=0}^{\infty} \frac{|a_k|^2}{k+1} \geqslant \pi|a_0|^2, \tag{3.70}$$

by a calculation like that at the beginning of the proof of Theorem 3.20. But

$$a_0 = \varphi(0) = \frac{F[f^{-1}(0)]}{F(z_0)}\,[f^{-1}(w)]'_{w=0} = \frac{1}{f'(z_0)},$$

since $f^{-1}(0) = z_0$, and hence

$$\iint_G \left|\frac{F(z)}{F(z_0)}\right|^2 dx\,dy \geqslant \frac{\pi}{|f'(z_0)|^2},$$

which is the inequality in (3.68). This inequality can become an equality only if

$$\pi|a_0|^2 = \pi \sum_{k=0}^{\infty} \frac{|a_k|^2}{k+1}$$

[cf. (3.70)], i.e., only if $a_1 = a_2 = \cdots = 0$. But then

$$\varphi(w) = a_0 = \frac{1}{f'(z_0)},$$

or equivalently,

$$\frac{F[f^{-1}(w)]}{F(z_0)} = \frac{1}{[f^{-1}(w)]'f'(z_0)},$$

which is just another way of writing (3.69).

We are now in a position to prove the following remarkable result:

THEOREM 3.27. *Let G be a Carathéodory domain, and let $\{P_k(z)\}$ be the polynomials orthonormal on G. Then the function $w = f(z)$ mapping G*

onto the unit disk $|w| < 1$ *and satisfying the conditions* $f(z_0) = 0$, $f'(z_0) > 0$ *is given explicitly by the expansion*

$$f(z) = \int_0^z f'(z)\,dz = \sqrt{\frac{\pi}{\sum_{k=0}^{\infty} |P_k(z_0)|^2}} \sum_{k=0}^{\infty} \overline{P_k(z_0)} \int_0^z P_k(z)\,dz. \quad (3.71)$$

Proof. Consider the polynomial

$$K_n(z, z_0) = \sum_{k=0}^{n} \overline{P_k(z_0)} P_k(z).$$

Obviously

$$K_n(z_0, z_0) = \sum_{k=0}^{n} |P_k(z_0)|^2 \neq 0,$$

since $P_0(z_0) \neq 0$, and hence, by (3.68),

$$\iint_G \left| \frac{K_n(z, z_0)}{K_n(z_0, z_0)} \right|^2 dx\,dy = \frac{\sum_{k=0}^{\infty} |P_k(z_0)|^2}{\left(\sum_{k=0}^{\infty} |P_k(z_0)|^2 \right)^2} = \frac{1}{\sum_{k=0}^{n} |P_k(z_0)|^2} \geqslant \frac{\pi}{|f'(z_0)|^2}.$$

Since this inequality holds for arbitrary n, the series

$$\sum_{k=0}^{\infty} |P_k(z_0)|^2$$

converges, and moreover

$$\frac{1}{\sum_{k=0}^{\infty} |P_k(z_0)|^2} \geqslant \frac{\pi}{|f'(z_0)|^2}.$$

It follows from Theorem 3.24 that

$$\sum_{k=0}^{\infty} \overline{P_k(z_0)} P_k(z)$$

is the Fourier series of a function $K(z, z_0) \in \mathscr{A}(G)$, where

$$\iint_G |K(z, z_0)|^2\,dx\,dy = \sum_{k=0}^{\infty} |P_k(z_0)|^2,$$

and hence

$$\iint_G \left| \frac{K(z, z_0)}{K(z_0, z_0)} \right|^2 dx\,dy = \frac{1}{\sum_{k=0}^{\infty} |P_k(z_0)|^2} \geqslant \frac{\pi}{|f'(z_0)|^2}. \quad (3.72)$$

Now let

$$\sum_{k=0}^{\infty} c_k P_k(z)$$

be the Fourier series of $f'(z)$. Then

$$\frac{\pi}{|f'(z_0)|^2} = \iint_G \left| \frac{f'(z)}{f'(z_0)} \right|^2 dx\, dy$$

$$= \frac{\sum_{k=0}^{\infty} |c_k|^2}{\left| \sum_{k=0}^{\infty} c_k P_k(z_0) \right|^2} \geqslant \frac{1}{\sum_{k=0}^{\infty} |P_k(z_0)|^2}, \tag{3.73}$$

where in the last step we use the Cauchy-Schwarz inequality

$$\left| \sum_{k=0}^{\infty} c_k P_k(z_0) \right|^2 \leqslant \sum_{k=0}^{\infty} |c_k|^2 \sum_{k=0}^{\infty} |P_k(z_0)|^2$$

(cf. footnote 26, p. 118). Comparing (3.72) and (3.73), we conclude that

$$\frac{\pi}{|f'(z_0)|^2} = \frac{1}{\sum_{k=0}^{\infty} |P_k(z_0)|^2},$$

i.e.,

$$f'(z_0) = \sqrt{\pi \sum_{k=0}^{\infty} |P_k(z_0)|^2}, \tag{3.74}$$

and moreover that the equality in (3.68) is achieved for the function

$$F(z) = K(z, z_0).$$

Then, according to (3.69),

$$\frac{K(z, z_0)}{K(z_0, z_0)} = \frac{f'(z)}{f'(z_0)}.$$

It follows that

$$f'(z) = \frac{f'(z_0)}{K(z_0, z_0)} K(z, z_0) = \sqrt{\frac{\pi}{\sum_{k=0}^{\infty} |P_k(z_0)|^2}} \sum_{k=0}^{\infty} \overline{P_k(z_0)} P_k(z).$$

Integrating this series term by term, we obtain (3.71), as required. It should be noted that the polynomials figuring in the uniformly convergent series (3.71) are no longer orthogonal.

Remark. The derivative of the function

$$\frac{f(z)}{f'(z_0)}$$

at the point z_0 equals 1, and hence this function maps G conformally onto the disk whose radius

$$R_{z_0} = \frac{1}{f'(z_0)}$$

is the conformal radius of the domain G relative to the point z_0 (see p. 12). It follows from (3.74) that

$$R_{z_0} = \frac{1}{\sqrt{\pi \sum_{k=0}^{\infty} |P_k(z_0)|^2}}.$$

We conclude our study of approximation theory by noting that for a domain bounded by a closed rectifiable Jordan curve Γ, there exists a well-developed theory of polynomials $\{p_n(z)\}$ orthonormal on Γ, in the sense that

$$\int_\Gamma p_m(z)\overline{p_n(z)}\, ds = \delta_{mn} \qquad (ds = |dz|),$$

of series expansions with respect to these polynomials, and so on.[30]

PROBLEMS

3.1. Given an open set $\mathcal{O}$, let $\{\mathcal{O}_n\}$ be the increasing sequence of open sets approximating $\mathcal{O}$ constructed on p. 81. Show that $\{\bar{\mathcal{O}}_n\}$ is an increasing sequence of compact sets approximating $\mathcal{O}$ in the sense of p. I.373.

3.2. What is meant by equation (3.2) if the component $\mathcal{C}$ is multiply connected?

3.3. Prove Theorem 3.2 by using Vitali's theorem, as in the proof of Theorem I.17.19.

3.4. Reconcile Theorem 3.3 with Problem 2.7.

3.5. Given a bounded domain G, show that there is a function analytic on G but not on any larger domain.

Hint. Let $\{\zeta_k\}$ be a sequence of points in G with no limit points in G which is dense in the boundary of G.[31] Then use Theorem 3.3.

---

[30] For a discussion of this theory, due to V. I. Smirnov, M. V. Keldysh, M. A. Lavrentev and P. P. Korovkin, among others, see V. I. Smirnov and N. A. Lebedev, *Constructive Theory of Functions of a Complex Variable* (in Russian), Izd. Nauka, Moscow (1964), Chap. 4. This book also contains a detailed treatment of Faber polynomials (Chap. 2), polynomials orthogonal on a domain (Chap. 3), and other topics in approximation theory.

[31] I.e., every point of the boundary is a limit point of $\{\zeta_k\}$.

3.6. Let $f(z)$ be a function meromorphic on the disk $K: |z| < R$, with poles $a_1, a_2, \ldots$ of orders $\alpha_1, \alpha_2, \ldots$ Then $f(z)$ is said to be of *bounded characteristic* if

$$\overline{\lim_{\rho \to R}} \frac{1}{2\pi} \int_0^{2\pi} \ln^+ |f(\rho e^{i\varphi})|\, d\varphi + \sum_{|a_k| < \rho} \alpha_k \ln \frac{\rho}{|a_k|} < \infty$$

[this condition generalizes (II.8.29)]. Let N be the class of all such functions. Prove the following generalization of Theorem II.8.7, also due to Nevanlinna: The class N coincides with the class of all functions meromorphic on K which can be represented in the form

$$f(z) = \frac{g_1(z)}{g_2(z)},$$

where $g_1(z)$ and $g_2(z)$ are analytic and bounded on K.[32] Give an example of a function which is meromorphic but not of bounded characteristic on the unit disk $|z| < 1$.

3.7. Let $\mathcal{O}$ and $\mathcal{O}'$ be the same as on p. 95. Prove that

a) $\mathcal{O} = \mathcal{O}'$ if and only if $\mathcal{O}$ is a union of disjoint simply connected domains;
b) $\mathcal{O}'$ is a union of disjoint simply connected domains;
c) Given any compact set $F \subset \mathcal{O}$, there is a closed Jordan curve $\Gamma \subset \mathcal{O}$ such that $F \subset I(\Gamma)$.

Hint. Recall the proof of Theorem I.4.14.

3.8. With the same notation as in the preceding problem, prove that

$$\mathcal{O}' = \bigcup_{\mathcal{C} \subset \mathcal{O}} \operatorname{int} \left(\bigcap_{G \supset \mathcal{C}} G \right),$$

where $\mathcal{C}$ is any component of $\mathcal{O}$, G is any simply connected domain, and the symbol int has the same meaning as in Problem I.4.14.

Comment. In this sense, $\mathcal{O}'$ is the smallest union of simply connected domains containing $\mathcal{O}$.

3.9. Justify the example on p. 96 in view of the fact that the lemma on p. 83 was proved only for finite ξ.

3.10. Show that the polynomial $P(z)$ figuring in Walsh's theorem (Theorem 3.9) can always be chosen to satisfy an auxiliary condition $P(\zeta) = f(\zeta)$, where ζ is any point in $\bar{G}$.

3.11. Show that if Γ is a piecewise smooth closed Jordan curve, than $I(\Gamma)$ can be represented to within finitely many domains of arbitrarily small diameter as a union of starlike domains (in the sense of Problem 1.15).

Hint. The behavior of Γ near cusps requires special investigation.

3.12. Give an elementary proof of Theorem 3.10 for the special case where Γ is a piecewise smooth closed Jordan curve.

Hint. Use Theorem I.13.3 and the result of the preceding problem.

---

[32] For the details, see G. M. Goluzin, *op. cit.*, Chap. 7, Sec. 5.

3.13. Let G be the domain shown in Figure 2.4, p. 38, and let

$$f(z) = \frac{1}{z - z_0},$$

where z_0 belongs to the inner disk-shaped domain. Is there a sequence of polynomials converging uniformly to $f(z)$

a) Inside G; b) On G; c) On $\overline{G}$?

How about nonuniform convergence?

3.14. With the notation of Sec. 14, prove that the continua F and $\overline{I(C_R)}$, $R > \rho$ generate the same Faber polynomials.

3.15. Prove that the zeros of the Chebyshev polynomials are all real and lie in the interval $-1 < x < 1$.

3.16. Prove that the Chebyshev polynomials obey the recurrence relation

$$T_{n+1}(z) = zT_n(z) - \tfrac{1}{4}T_{n-1}(z).$$

3.17. Prove that the Chebyshev polynomials satisfy the orthogonality relation

$$\int_{-1}^{1} \frac{T_m(x)T_n(x)}{\sqrt{1 - x^2}}\, dx = 0 \qquad (m \neq n).$$

3.18. Prove that among all polynomials

$$P(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$$

of degree $n \geqslant 1$, with leading coefficient 1, the Chebyshev polynomial $T_n(x)$ deviates least from zero in the interval $-1 \leqslant x \leqslant 1$, in the sense that it minimizes the expression

$$\max_{-1 \leqslant x \leqslant 1} |P(x)|.$$

3.19. Prove that among all polynomials

$$P(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_0$$

of degree $n \geqslant 1$, with leading coefficient 1, the polynomial $\Pi(z) = z^n$ minimizes the expression

$$\max_{|z| \leqslant R} |P(z)|. \tag{3.75}$$

3.20. Prove that among all polynomials of the form

$$P(z) = Az^n + a_{n-1}z^{n-1} + \cdots + a_1 z + B$$

of degree $n \geqslant 2$, with fixed coefficients A and B, the polynomial $\Pi(z) = Az^n + B$ minimizes the expression (3.75).

3.21. Prove that the limit figuring in footnote 25, p. 116 is independent of the choice of the increasing sequence $\{D_n\}$ approximating G.

3.22. Let $\{F_n(z)\}$ be a sequence of functions in $\mathscr{A}(G)$ converging to a function $F(z) \in \mathscr{A}(G)$. Prove that $F(z)$ is unique.

3.23. Let $\{F_n(z)\}$ be a sequence of functions in $\mathscr{A}(G)$ such that

$$\lim_{m,n\to\infty} \iint_G |F_m(z) - F_n(z)|^2 \, dx \, dy = 0.$$

Prove that $\{F_n(z)\}$ converges in the mean to a function $F(z) \in \mathscr{A}(G)$.

3.24. Find the sequence of orthonormal polynomials $\{P_n(z)\}$ satisfying Theorem 3.22 if G is

a) The unit disk $|z| < 1$;
b) The open square $-1 < x < 1$, $-1 < y < 1$;
c) The "slit disk" $|z| < 1$, $z \notin [0, 1)$.

3.25. Let G be the slit disk of the preceding problem. Then $f(z) = \sqrt{z}$ is analytic on G. Show that $f(z)$ is not the limit in the mean of its Fourier series. Reconcile this fact with Theorem 3.25.

3.26. Let $\Phi_n(z)$ be any of the orthonormal polynomials on G constructed in Theorem 3.22. Prove that the zeros of $\Phi_n(z)$ all lie in the smallest convex domain containing G.

Hint. If ζ is a zero of $\Phi_n(z)$, then

$$\frac{\Phi_n(z)}{z - \zeta}$$

is orthogonal to $\Phi_n(z)$.

Part 2

PERIODIC AND ELLIPTIC FUNCTIONS

4

PERIODIC MEROMORPHIC FUNCTIONS

18. Preliminaries

We begin with a definition anticipated on p. I.140:

DEFINITION 1. *A function $f(z)$ is said to be periodic if there exists a nonzero complex number ω, called a period of $f(z)$, such that*

$$f(z + \omega) = f(z) \tag{4.1}$$

for all z. The number 0 is also regarded as a period of $f(z)$. If z_0 is a pole of $f(z)$, so is $z_0 + \omega$.

Remark 1. Constant functions are trivially periodic.[1]

Remark 2. If ω is a period of $f(z)$, so is $-\omega$. In fact, replacing ω by $-\omega$ in (4.1), we obtain

$$f(z) = f(z - \omega)$$

for all z.

THEOREM 4.1. *If $\omega_1, \ldots, \omega_n$ are all periods of $f(z)$, then so is $m_1\omega_1 + \cdots + m_n\omega_n$, where $m_1, \ldots, m_n$ are arbitrary integers.*

Proof. If ω_1 and ω_2 are two periods of $f(z)$, distinct or not, then

$$f(z + \omega_1 + \omega_2) = f(z + \omega_1) = f(z),$$

[1] Every complex number is a period of a constant function.

and hence $\omega_1 + \omega_2$ is a period of $f(z)$. By induction, if $\omega_1, \ldots, \omega_n$ are all periods of $f(z)$, so is $\omega_1 + \cdots + \omega_n$. In particular, if ω is a period, so are

$$m\omega = \underbrace{\omega + \cdots + \omega}_{m \text{ times}}, \quad 0 \cdot \omega = 0, \quad -m\omega = \underbrace{-\omega - \cdots - \omega}_{m \text{ times}}$$

(recall Remark 2). The theorem now follows by another induction.

THEOREM 4.2. *Let E be the set of all periods of a periodic meromorphic function $f(z) \not\equiv$ const. Then E has no finite limit points.*

Proof. Suppose E has a finite limit point ω_0, and let $\{\omega_n\}$ be a sequence of distinct periods of $f(z)$ converging to ω_0. Then $f(z)$ takes the same value $f(0)$ at every point ω_n, since obviously

$$f(0) = f(0 + \omega_n) = f(\omega_n) \qquad (n = 1, 2, \ldots).$$

But this is incompatible with the assumption that $f(z)$ is meromorphic and nonconstant. In fact, $f(z)$ has a limit point of poles if $f(0) = \infty$, while $f(z) \equiv f(0)$ by the interior uniqueness theorem (Theorem I.17.1) if $f(0) \neq \infty$.

THEOREM 4.3. *Let $f(z) \not\equiv$ const be meromorphic and periodic, with a nonzero period ω_1'. Then the line L drawn through the points 0 and ω_1' contains a nonzero period ω_1 of $f(z)$ such that every period of $f(z)$ lying on L is of the form $m\omega_1$, where m is an integer.*

Proof. According to Theorem 4.2, the line segment δ' joining the points 0 and ω_1' can only contain finitely many periods of $f(z)$. Of these periods, let ω_1 be the nonzero period closest to the origin (ω_1 may coincide with ω_1'). Then δ, the line segment joining 0 and ω_1, contains no periods of $f(z)$ other than 0 and ω_1. Suppose $\Omega \in L$ is a period of $f(z)$ which is not an integral multiple of ω_1. Then Ω lies between two points $m\omega_1$ and $(m + 1)\omega_1$ on L (see Figure 4.1), and hence

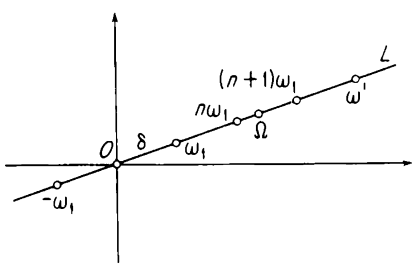


FIGURE 4.1

$$0 < |\Omega - m\omega_1| < |\omega_1|.$$

Therefore $\Omega - m\omega_1$ is a period of $f(z)$ contained in δ but distinct from 0 and ω_1, contrary to the construction of δ. This contradiction proves the theorem.

DEFINITION 2. *A periodic function $f(z)$ is said to be simply periodic if all its periods are of the form $m\omega_1$, where m is an arbitrary integer. The*

number ω_1 *(or* $-\omega_1$*) is called the fundamental period of* $f(z)$.[2] *Constant functions are regarded as simply periodic.*

Example. The function e^z is simply periodic, with fundamental period $2\pi i$ or $-2\pi i$ (see p. I.140).

THEOREM 4.4. *Let* $f(z) \not\equiv$ const *be meromorphic and periodic, with two nonzero periods* ω_1' *and* ω_2' *such that the points* 0, ω_1' *and* ω_2' *are noncollinear. Then* $f(z)$ *has two nonzero periods* ω_1 *and* ω_2 *such that every period of* $f(z)$ *is of the form* $m_1\omega_1 + m_2\omega_2$, *where* m_1 *and* m_2 *are integers.*

Proof. Let L and ω_1 be the same as in Theorem 4.3. Then ω_2' does not lie on L. Consider the closed triangular domain T' with vertices 0, ω_1 and ω_2' (see Figure 4.2), and let E be the set of periods of $f(z)$ which are contained in T' but are distinct from 0 and ω_1. Clearly E is a finite set, with no points on the line segment joining 0 and ω_1. Let ω_2 be the point of E closest to the line L (ω_2 may coincide with ω_2'). Then the closed triangular domain T with vertices 0, ω_1 and ω_2 (T may coincide with T') contains no periods of $f(z)$ other than its vertices. The parallelogram Δ with vertices 0, ω_1, ω_2 and $\omega_1 + \omega_2$ also contains no periods of $f(z)$ other than its vertices. In fact, suppose ω is a period of $f(z)$ distinct from 0, ω_1, ω_2, $\omega_1 + \omega_2$ and contained in Δ. Then $\omega^* = \omega_1 + \omega_2 - \omega$ is another period of $f(z)$ with the same properties,[3] and moreover either ω or ω^* belongs to T, contrary to the construction of T.

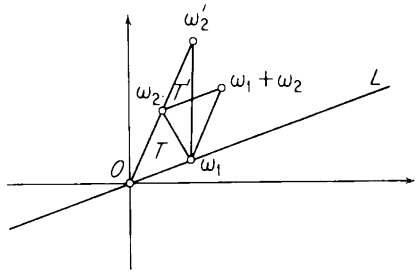


FIGURE 4.2

Now, given any two integers m_1 and m_2, let $\Delta_{m_1 m_2}$ be the parallelogram with vertices

$$m_1\omega_1 + m_2\omega_2, \quad (m_1 + 1)\omega_1 + m_2\omega_2,$$
$$m_1\omega_1 + (m_2 + 1)\omega_2, \quad (m_1 + 1)\omega_1 + (m_2 + 1)\omega_2.$$

Clearly $\Delta_{m_1 m_2}$ is congruent to $\Delta = \Delta_{00}$, and the set of all $\Delta_{m_1 m_2}$ for arbitrary integral m_1, m_2 makes up a "parallelogram lattice" in the z-plane, whose vertices consist of all points of the form $m_1\omega_1 + m_2\omega_2$ (see Figure 4.3). Suppose Ω is a period of $f(z)$ which is not of the form $m_1\omega_1 + m_2\omega_2$. Then Ω is contained in some parallelogram $\Delta_{m_1 m_2}$ and is distinct from the vertices of $\Delta_{m_1 m_2}$, as shown in the figure. Therefore $\Omega - m_1\omega_1 - m_2\omega_2$ is a period of $f(z)$ contained in Δ but distinct from

---

[2] The synonymous term *primitive period* is also widely used (similarly in Definition 3 below).

[3] Note that ω^* is the reflection of ω in the point $\frac{1}{2}(\omega_1 + \omega_2)$.

the vertices of Δ, contrary to the construction of Δ. This contradiction shows that Ω must coincide with one of the vertices of the lattice, thereby proving the theorem.

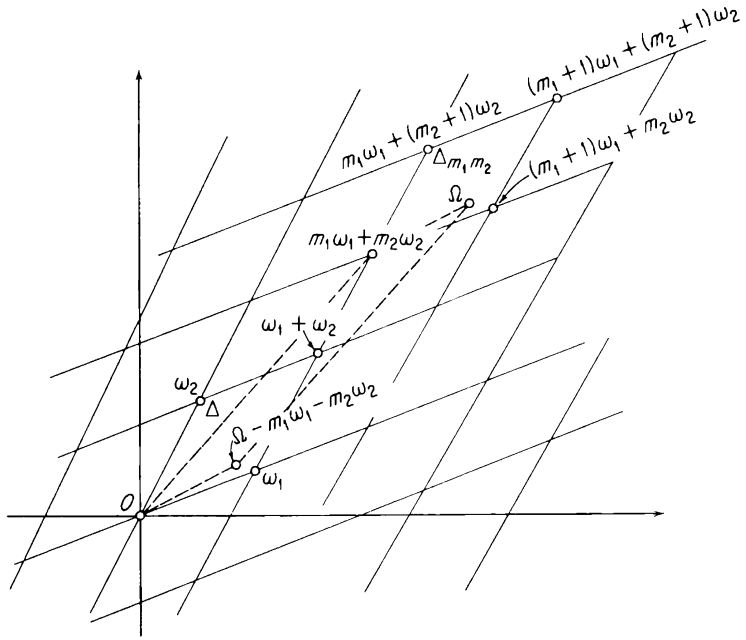


FIGURE 4.3

DEFINITION 3. *A periodic function $f(z)$ is said to be doubly periodic if all its periods are of the form $m_1\omega_1 + m_2\omega_2$, where m_1 and m_2 are arbitrary integers and the points 0, ω_1, ω_2 are noncollinear. The numbers ω_1 and ω_2 are called fundamental periods of $f(z)$, and the parallelogram with vertices 0, ω_1, ω_2, $\omega_1 + \omega_2$ is called a fundamental period parallelogram. Constant functions are regarded as doubly periodic.*

DEFINITION 4. *A function is said to be elliptic if it is doubly periodic and meromorphic.*

THEOREM 4.5. *A periodic meromorphic function is either simply periodic or doubly periodic.*

Proof. Combine Theorems 4.3 and 4.4.

THEOREM 4.6. *Given a doubly periodic function $f(z)$ with fundamental periods ω_1 and ω_2, let ω'* and *ω'' be two periods of $f(z)$ of the form*

$$\begin{aligned} \omega' &= m_1'\omega_1 + m_2''\omega_2, \\ \omega'' &= m_1''\omega_1 + m_2''\omega_2, \end{aligned} \tag{4.2}$$

where

$$m_1'm_2'' - m_1''m_2' = \pm 1. \tag{4.3}$$

Then ω' and ω'' are also fundamental periods of $f(z)$.

Proof. Solving (4.2) for ω_1 and ω_2, we obtain

$$\omega_1 = \pm(m_2''\omega' - m_2'\omega''),$$
$$\omega_2 = \pm(-m_1''\omega' + m_1''\omega'').$$

Therefore any period $\Omega = m_1\omega_1 + m_2\omega_2$ can be written as a linear combination of ω' and ω'' with integral coefficients:

$$\Omega = \pm[(m_1m_2'' - m_2m_1')\omega' + (-m_1m_2' + m_2m_1')\omega''].$$

In other words, ω' and ω'' are a pair of fundamental periods of $f(z)$, as asserted.

Remark. Thus $f(z)$ has infinitely many pairs of fundamental periods, since there are infinitely many sets of four integers satisfying (4.3). Similarly, $f(z)$ has infinitely many fundamental period parallelograms.

The next two theorems are valid for any periodic meromorphic function, whether simply or doubly periodic. Suppose $f(z)$ is periodic, with period $\omega \neq 0$. Then there is no loss of generality in assuming that $\omega = 2\pi$. In fact, given a periodic function $F(\zeta)$ with nonzero period $\omega \neq 2\pi$, the preliminary substitution

$$z = \frac{2\pi}{\omega}\zeta \tag{4.4}$$

transforms $F(\zeta)$ into a periodic function

$$f(z) = F\left(\frac{\omega}{2\pi}z\right)$$

with period 2π (interpret this transformation geometrically).

THEOREM 4.7. *If $f(z)$ is meromorphic and periodic with period 2π, then*

$$f^*(t) = f\left(\frac{1}{i}\operatorname{Ln} t\right) \tag{4.5}$$

is meromorphic on the domain G equal to the finite t-plane minus the point $t = 0$. The poles of $f^*(t)$ coincide with those of $f(z)$ and are of the same order.

Proof. At every point $t \in G$, the function

$$z = \frac{1}{i}\operatorname{Ln} t$$

has a countably infinite set of values, differing by integral multiples of 2π. But $f(z)$ is periodic with period 2π. Therefore the function $f^*(t)$ is

single-valued. Let z_0 be any value of $(1/i)$ Ln t, let $\mathcal{N}(t_0) \subset G$ and let $\varphi_0(t)$ be the single-valued analytic branch of $(1/i)$ Ln t equal to z_0 at $t = t_0$. Clearly $\varphi_0[\mathcal{N}(t_0)]$ is a domain in the z-plane containing the point z_0. If z_0 is a regular point of $f(z)$, we can use the continuity of $\varphi_0(t)$ to choose $\mathcal{N}(t_0)$ so small that the domain $\varphi_0[\mathcal{N}(t_0)]$ contains only regular points of $f(z)$. Then $f^*(t) = f[\varphi_0(t)]$ is analytic on $\mathcal{N}(t_0)$, being an analytic function of an analytic function. If z_0 is a pole of $f(z)$, we can again choose $\mathcal{N}(t_0)$ so small that z_0 is the only pole of $f(z)$ in $\varphi_0[\mathcal{N}(t_0)]$. Therefore, in this case $f^*(t) = f[\varphi_0(t)]$ will be single-valued and analytic on $\mathcal{N}(t_0)$ except at the point $t = t_0$. But

$$\lim_{t \to t_0} f^*(t) = \lim_{z \to z_0} f(z) = \infty,$$

and hence t_0 is a pole of $f^*(t)$. Suppose $f(z)$ has a pole of order k at z_0, so that

$$\lim_{z \to z_0} (z - z_0)^k f(z) = A,$$

where $A = 0, \infty$. Then

$$\lim_{t \to t_0} (t - t_0)^k f^*(t) = \lim_{t \to t_0} \frac{(z - z_0)^k f(z)}{\left(\dfrac{z - z_0}{t - t_0}\right)^k}$$

$$= \lim_{t \to t_0} \frac{(z - z_0)^k f(z)}{\left[\dfrac{\varphi(t) - \varphi(t_0)}{t - t_0}\right]^k} = \frac{A}{[\varphi'(t_0)]^k},$$

where the last expression cannot equal 0 or ∞, since

$$\varphi'(t_0) = \frac{d}{dt}\left(\frac{1}{i} \operatorname{Ln} t\right)\Big|_{t=t_0} = \frac{1}{it_0} \neq 0, \infty.$$

In other words, $f^*(t)$ has a pole of order k at t_0. Since the only singular points of $f^*(t)$ in G are poles (why?), the proof is complete.

An obvious modification of this proof leads to the converse of Theorem 4.7:

THEOREM 4.7′. *If $f^*(t)$ is meromorphic on the domain G equal to the finite t-plane minus the point $t = 0$, then*

$$f(z) = f^*(e^{iz}) \tag{4.6}$$

is meromorphic (in the finite z-plane) and periodic with period 2π.

Remark. According to Theorems 4.7 and 4.7′, there is a one-to-one correspondence [established by (4.5) and (4.6)] between periodic meromorphic functions $f(z)$ of period 2π and functions $f^*(t)$ meromorphic on G.

Let $f(z)$ be periodic with fundamental period 2π. Then every open strip σ of width 2π bounded by two lines parallel to the imaginary axis is called a

period strip of $f(z)$. As we know from Sec. I.39, the function $t = e^{iz}$ maps σ conformally onto a domain Σ equal to the finite t-plane minus a ray Λ joining the points 0 and ∞,[4] where Λ is the image of both of the lines forming the boundary of σ. If $f(z)$, $f^*(t)$ and G are the same as in Theorems 4.7 and 4.7′, then clearly $G = \Sigma \cup \Lambda$, and hence the values taken by $f^*(t)$ in G coincide with those taken by $f(z)$ in the half-closed period strip $g = \sigma \cup \lambda$, where λ is either of the two lines bounding σ. In particular, $f^*(t)$ has the same number of poles in G as $f(z)$ has in g.

19. Periodic Entire Functions. Trigonometric Polynomials

THEOREM 4.8. *The class of periodic entire functions of period* 2π *coincides with the class of functions which have Fourier expansions of the form*

$$f(z) = \sum_{k=-\infty}^{\infty} a_k e^{-ikz}. \tag{4.7}$$

Here

$$a_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx}\, dx \qquad (k = 0, \pm 1, \pm 2, \ldots), \tag{4.8}$$

and the series (4.7) *is uniformly convergent on every horizontal strip containing the real axis.*

Proof. The fact that (4.7) is a periodic entire function of period 2π is obvious [by hypothesis, (4.7) converges on every compact set]. Conversely, given any periodic entire function $f(z)$, let $f^*(t)$ and G be the same as in Theorem 4.7. Then $f^*(t)$ has no poles in G, and hence has a Laurent expansion of the form

$$f^*(t) = \sum_{k=-\infty}^{\infty} a_k t^k \qquad (t \neq 0, \infty), \tag{4.9}$$

where the convergence is uniform on every annulus $0 < r \leqslant |t| \leqslant R < \infty$. Therefore the series

$$f(z) = \sum_{k=-\infty}^{\infty} a_k e^{ikz} \tag{4.10}$$

is uniformly convergent on every strip $\ln r \leqslant -y \leqslant \ln R$ or

$$\ln \frac{1}{R} \leqslant y \leqslant \ln \frac{1}{r} \qquad (0 < r < R < \infty),$$

i.e., on every horizontal strip with sides parallel to the real axis. In particular, (4.10) is uniformly convergent on the real axis itself. Multiplying

[4] Since σ lies in the finite z-plane, Λ is open, i.e., Λ does not contain its end points 0 and ∞.

(4.10) by e^{-inz} and then integrating term by term along the segment $\delta: -\pi \leqslant x \leqslant \pi$, $y = 0$, we find that

$$\int_\delta f(z)e^{-inz}\,dz = \int_{-\pi}^{\pi} f(x)e^{-inx}\,dx = \sum_{k=-\infty}^{\infty} a_k \int_{-\pi}^{\pi} e^{i(k-n)x}\,dx = 2\pi a_n, \tag{4.11}$$

because of the orthogonality of the functions e^{ikx} ($k = 0, 1, 2, \ldots$) on the interval $[-\pi, \pi]$. But (4.11) implies (4.8), as required.

COROLLARY. *The expansion* (4.7) *can be written in the form*

$$f(z) = \frac{\alpha_0}{2} + \sum_{k=1}^{\infty} (\alpha_k \cos kz + \beta_k \sin kz), \tag{4.12}$$

where

$$\alpha_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos kx\,dx, \quad \beta_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin kx\,dx \qquad (k = 0, 1, 2, \ldots). \tag{4.13}$$

Proof. First note that

$$f(z) = a_0 + \sum_{k=1}^{\infty} [(a_k + a_{-k})\cos kz + i(a_k - a_{-k})\sin kz],$$

where

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,dx,$$

$$a_k + a_{-k} = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\frac{e^{-ikx} + e^{ikx}}{2}\,dx = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos kx\,dx,$$

$$i(a_k - a_{-k}) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\frac{e^{ikx} - e^{-ikx}}{2i}\,dx = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin kx\,dx,$$

and then write

$$a_0 = \frac{\alpha_0}{2}, \quad a_k + a_{-k} = \alpha_k, \qquad i(a_k - a_{-k}) = \beta_k.$$

Remark. Naturally, the series (4.12) is again uniformly convergent on every horizontal strip containing the real axis.

A periodic entire function of the special form

$$f(z) = \sum_{k=-m}^{m} a_k e^{ikz}, \tag{4.14}$$

where a_m and a_{-m} do not both vanish, is called a *trigonometric polynomial of degree m* (and period 2π). Such functions have already been encountered in Sec. I.52. We can also write

$$f(z) = \frac{\alpha_0}{2} + \sum_{k=1}^{\infty} (\alpha_k \cos kz + \beta_k \sin kz),$$

where the coefficients are given by (4.13).

THEOREM 4.9. *The class of trigonometric polynomials coincides with the class of periodic entire functions $f(z)$ which approaches definite limits (finite or infinite) as z approaches the upper and lower ends of a period strip.*

Proof. Every trigonometric polynomial obviously has the property in question. Conversely, if $f(z)$ is an entire function with the given property, then each of the points $t = 0$ and $t = \infty$ is a regular point or a pole of $f^*(t)$, and not an essential singular point (cf. Problems 4.11–4.13). Therefore the Laurent expansion (4.9) reduces to

$$f^*(t) = \sum_{k=-m}^{m} a_k t^k$$

(either a_m or a_{-m} may vanish), and hence $f(z)$ is of the form (4.14), as asserted.

THEOREM 4.10. *A periodic entire function $f(z) \not\equiv$ const is a trigonometric polynomial if and only if it is a function of exponential type. Then $f(z)$ has order 1 and type m equal to the degree of the polynomial.*

Proof. Suppose $f(z)$ is a trigonometric polynomial of degree m. Then clearly

$$|f(z)| < e^{(m+\varepsilon)|z|} \qquad (\varepsilon > 0)$$

for all sufficiently large $|z|$, while on the other hand there exist numbers z_n of arbitrarily large modulus such that

$$|f(z)| > e^{(m-\varepsilon)|z_n|} \qquad (\varepsilon > 0)$$

(at least one of the numbers a_m, a_{-m} is nonzero). Therefore $f(z)$ is of order 1 and type m (see p. II.251), and hence of exponential type (see p. II.260).

Conversely, suppose $f(z)$ is of exponential type with period 2π. Then

$$|f(z)| < e^{C|z|}$$

for some C $(0 < C < \infty)$ and all sufficiently large $|z|$. Therefore on the line segments $y = \pm \ln R$ inside the period strip $-\pi \leqslant x \leqslant \pi$, we find that

$$|f(x + iy)| < \exp\{C|x \pm i \ln R|\} = \exp\left\{C \ln R \left|1 \pm i\frac{x}{\ln R}\right|\right\}$$

$$< \exp\left\{C \ln R\left(1 + \frac{\pi}{\ln R}\right)\right\} < \exp\{(C + \varepsilon) \ln R\} = R^{C+\varepsilon}$$

for all $R > R(\varepsilon)$. Under the transformation $t = e^{iz}$, the line segments in question go into the circles $|t| = 1/R$ and $|t| = R$ in the domain G figuring in Theorem 4.7. Therefore

$$|f^*(t)| < \frac{1}{|t|^{C+\varepsilon}} \tag{4.15}$$

on the first of the circles, while

$$|f^*(t)| < |t|^{C+\varepsilon} \tag{4.16}$$

on the second, provided that R is sufficiently large. Thus (4.15) holds on all circles $|t| = \rho$ of sufficiently small radius

$$\rho = \frac{1}{R} < \frac{1}{R(\varepsilon)},$$

while (4.16) holds on all circles $|t| = \rho$ of sufficiently large radius

$$\rho = R > R(\varepsilon).$$

If a_{-k} is the coefficient of t^{-k} in the Laurent series (4.9), then

$$|a_{-k}| = \left| \frac{1}{2\pi i} \int_{|t|=\rho} \frac{f^*(t)}{t^{-k+1}}\, dt \right| = \frac{\rho^k}{2\pi} \left| \int_0^{2\pi} f^*(e^{i\theta}) e^{ik\theta}\, d\theta \right|,$$

and hence by (4.15),

$$|a_{-k}| < \frac{\rho^k}{2\pi} \frac{2\pi}{\rho^{C+\varepsilon}} = \rho^{k-(C+\varepsilon)}$$

for all sufficiently small ρ. Letting $\rho \to 0$, we find that

$$a_{-k} = 0 \quad \text{if} \quad k > C + \varepsilon,$$

which implies

$$a_{-k} = 0 \quad \text{if} \quad k > [C],$$

since $\varepsilon > 0$ is arbitrary.[5] Similarly, if a_k is the coefficient of t^k in (4.9), we have

$$|a_k| = \frac{1}{2\pi i} \left| \int_{|t|=\rho} \frac{f^*(t)}{t^{k+1}}\, dt \right| = \frac{1}{2\pi\rho^k} \left| \int_0^{2\pi} f^*(\rho e^{i\theta}) e^{ik\theta}\, d\theta \right|,$$

and hence by (4.16),

$$|a_k| < \frac{1}{2\pi\rho^k} 2\pi\rho^{C+\varepsilon} = \rho^{(C+\varepsilon)-k}.$$

Letting $\rho \to \infty$, we obtain

$$a_k = 0 \quad \text{if} \quad k > C + \varepsilon$$

or

$$a_k = 0 \quad \text{if} \quad k > [C].$$

It follows that

$$f^*(t) = \sum_{k=-[C]}^{[C]} a_k t^k,$$

[5] $[C]$ denotes the integral part of C, i.e., the largest integer $\leqslant C$.

which implies

$$f(z) = \sum_{k=-[C]}^{[C]} a_k e^{ikz},$$

i.e., $f(z)$ is a trigonometric polynomial, as asserted. Clearly $f(z)$ is of degree $[C]$, order 1 and type $[C]$.

THEOREM 4.11. *The class of periodic meromorphic functions of period* 2π *coincides with the class of functions of the form*

$$f(z) = \frac{g(z)}{h(z)}, \tag{4.17}$$

where $g(z)$ *and* $h(z) \not\equiv 0$ *are periodic entire functions of period* 2π.

Proof. Let $f(z)$ be a function of the form (4.17), where $g(z)$ and $h(z) \not\equiv 0$ are entire and periodic with period 2π. Then $f(z)$ is obviously meromorphic and periodic with period 2π. Conversely, let $f(z)$ be any periodic meromorphic function of period 2π. Being meromorphic, $f(z)$ must have a representation of the form (4.17), where $g(z)$ and $h(z)$ are entire, but the fact that $g(z)$ and $h(z)$ are periodic requires further proof.[6] Let $f^*(t)$ and G be the same as in Theorem 4.6. Then $f^*(t)$ is meromorphic on G, and hence by Theorem 3.4,

$$f^*(t) = \frac{g^*(t)}{h^*(t)},$$

where $g^*(t)$ and $h^*(t) \not\equiv 0$ are analytic on G. It follows from Theorem 4.7′ that

$$f(z) = \frac{g^*(e^{iz})}{h^*(e^{iz})} = \frac{g(z)}{h(z)},$$

where the functions

$$g(z) = g^*(e^{iz}), \qquad h(z) = h^*(e^{iz}) \not\equiv 0$$

are obviously entire and periodic with period 2π. The proof is now complete.

COROLLARY. *Equation* (4.17) *can be written in the form*

$$f(z) = \frac{\sum_{k=-\infty}^{\infty} a_k e^{ikz}}{\sum_{k=-\infty}^{\infty} b_k e^{ikz}},$$

---

[6] For example,

$$\sin z = \frac{z \sin z}{z}$$

where $z \sin z$ and z are entire but not periodic.

or as a ratio of two convergent Fourier series:

$$f(z) = \frac{\frac{\alpha_0}{2} + \sum_{k=1}^{\infty} (\alpha_k \cos kz + \beta_k \sin kz)}{\frac{\alpha_0'}{2} + \sum_{k=1}^{\infty} (\alpha_k' \cos kz + \beta_k' \sin kz)}. \tag{4.18}$$

Remark 1. Suppose $f(z)$ has only finitely many poles in every period strip g (see p. 141) and approaches definite limits (finite or infinite) as z approaches the upper and lower ends of g. Then $f^*(t)$ has only finitely many poles in G, and each of the points $t = 0$ and $t = \infty$ is a regular point or a pole of $f^*(t)$ [cf. Problems 4.11–4.13]. Therefore $f^*(t)$ has no singular points other than poles in the *extended* t-plane, and hence is of the form

$$f^*(t) = \frac{a_0 + a_1 t + \cdots + a_m t^m}{b_0 + b_1 t + \cdots + b_n t^n}.$$

It follows that

$$f(z) = \frac{a_0 + a_1 e^{iz} + \cdots + a_m e^{imz}}{b_0 + b_1 e^{iz} + \cdots + b_n e^{inz}}, \tag{4.19}$$

i.e., $f(z)$ is a trigonometric function (in the sense of Sec. I.52). Conversely, every function of the form (4.19) has only finitely many poles in g and approaches definite limits as z approaches the upper and lower ends of g.

Remark 2. If $f(z)$ is periodic with nonzero period $\omega \neq 2\pi$, we need only replace z by $2\pi z/\omega$ in the right-hand sides of (4.12) and (4.18) [cf. (4.4)].

20. Elliptic Functions

Before proceeding, we need a somewhat sharper notation than in the proof of Theorem 4.4 and the subsequent definition. Let $f(z)$ be an elliptic function with fundamental periods ω', ω'' and fundamental period parallelogram Δ.[7] If $\operatorname{Im}(\omega''/\omega') > 0$ we write

$$\omega' = 2\omega_1, \qquad \omega'' = 2\omega_3,$$

while if $\operatorname{Im}(\omega'/\omega'') > 0$ we write[8]

$$\omega' = 2\omega_3, \qquad \omega'' = 2\omega_1.$$

Moreover, we denote the complex number $2\omega_1 + 2\omega_3$ (representing a diagonal of Δ) by $2\omega_2$. As a result, the vertices of Δ are labelled 0, $2\omega_1$, $2\omega_2$ and $2\omega_3$ when read in counterclockwise order, i.e., as encountered by an

[7] As in the remark on p. 139, ω', ω'' and Δ are not unique.

[8] Since 0, ω' and ω'' are noncollinear, either $\operatorname{Im}(\omega''/\omega') > 0$ or $\operatorname{Im}(\omega'/\omega'') > 0$.

observer traversing the boundary of Δ with the interior of Δ always on the left. Let $[a, b]$ be the line segment joining a to b, with equation

$$z = (1 - t)a + tb \qquad (0 \leqslant t \leqslant 1)$$

(thereby generalizing the notation for a closed interval), and consider the closed Jordan curve

$$\Gamma = [0, 2\omega_1] \cup [2\omega_1, 2\omega_2] \cup [2\omega_2, 2\omega_3] \cup [2\omega_3, 0]$$

(orientation is unimportant here). Then by the fundamental period parallelogram Δ, we henceforth mean the half-closed domain

$$\Delta = \overline{I(\Gamma)} - [2\omega_1, 2\omega_2] - [2\omega_2, 2\omega_3],$$

i.e., the closed parallelogram with vertices 0, $2\omega_1$, $2\omega_2$ and $2\omega_3$ minus the two sides sharing the point $2\omega_2$.[9] Similarly, let Δ_{mn} denote the parallelogram obtained by subjecting $\Delta = \Delta_{00}$ to the shift $2m\omega_1 + 2n\omega_3$ (thus $z \in \Delta_{mn}$ if and only if $z - 2m\omega_1 - 2n\omega_3 \in \Delta$), so that Δ_{mn} has the vertices

$$2m\omega_1 + 2n\omega_3, \quad (2m + 1)\omega_1 + 2n\omega_3,$$
$$(2m + 1)\omega_1 + (2n + 1)\omega_3, \quad 2m\omega_1 + (2n + 1)\omega_3$$

(read in counterclockwise order) and does not contain the two sides sharing the point

$$2m\omega_1 + 2n\omega_3 + 2\omega_2 = (2m + 1)\omega_1 + (2n + 1)\omega_3.$$

DEFINITION 1. *Two points z' and z'' in the finite plane are said to be congruent (relative to the fundamental periods $2\omega_1$ and $2\omega_3$) if $z'' - z'$ is a period, i.e., if $z'' - z' = 2m\omega_1 + 2n\omega_3$ where m and n are integers.*

The new notation, together with this definition, has the following obvious implications which are worth stating as theorems:

THEOREM 4.12. *The parallelograms Δ_{mn} and $\Delta_{m'n'}$ are disjoint unless $m = m'$, $n = n'$. Moreover, the finite plane is the union of all the parallelograms Δ_{mn} $(m, n = 0, \pm 1, \pm 2, \ldots)$.*

THEOREM 4.13. *No two distinct points of a parallelogram Δ_{mn} are congruent. Moreover, given any point z in the finite plane and any parallelogram Δ_{mn}, there is a unique point in Δ_{mn} congruent to z.*

We now prove a number of general properties of elliptic functions as a prelude to the actual construction of such functions.

---

[9] Another way of writing Δ is

$$\Delta = I(\Gamma) \cup [0, 2\omega_1) \cup [0, 2\omega_3),$$

where $[a, b)$ denotes the *half-open* line segment joining a to b, with equation

$$z = (1 - t)a + tb \qquad (0 \leqslant t < 1).$$

THEOREM 4.14. *The class of elliptic functions with fundamental periods* $2\omega_1$ *and* $2\omega_3$ *is closed under addition, subtraction, multiplication, division (by nonzero divisors) and differentiation.*

Proof. The class of meromorphic functions is closed under the operations in question, and so is the class of periodic functions. For example, if

$$f(z + 2m\omega_1 + 2n\omega_3) = f(z)$$

for all z, then

$$f'(z + 2m\omega_1 + 2n\omega_3) = f'(z)$$

for all z, and so on.

THEOREM 4.15. *If* $f(z)$ *is an entire elliptic function, then* $f(z) \equiv$ const.

Proof. Being entire, $f(z)$ is continuous and hence bounded on the closed parallelogram $\bar{\Delta}$. But according to Theorem 4.13, the value of $f(z)$ at any point of the finite plane is the same as its value at some point of Δ. Therefore $|f(z)| \leqslant M < \infty$ for all finite z, and hence $f(z) \equiv$ const by Liouville's theorem (Theorem I.16.8).

COROLLARY. *An elliptic function* $f(z) \not\equiv$ const *must have at least one pole in* Δ. *Moreover, the total number of poles in* Δ *must be finite.*

Proof. Use Theorem 4.15, and note that if Δ contains infinitely many poles, then Δ must contain a limit point of poles, contrary to the assumption that $f(z)$ is meromorphic.

DEFINITION 2. *The total number of poles of an elliptic function* $f(z)$ *in* Δ, *each counted a number of times equal to its order, is called the order of* $f(z)$.[10]

THEOREM 4.16. *The sum of the residues of an elliptic function* $f(z)$ *at all its poles in* Δ *equals zero.*

Proof. Let ρ be the distance between the set of poles of $f(z)$ in Δ and the set $[2\omega_1, 2\omega_2] \cup [2\omega_2, 2\omega_3]$, where $\rho = 0$ if Γ, the boundary of Δ, contains no poles. Then, shifting Δ along the diagonal $[2\omega_2, 0]$ by an amount less than ρ (the shift is in the direction from $2\omega_2$ to 0, and there is no shift if $\rho = 0$), we obtain a new parallelogram Δ' with vertices

$$z_0, z_0 + 2\omega_1, \quad z_0 + 2\omega_2, \quad z_0 + 2\omega_3$$

($z_0 = 0$ if $\rho = 0$). The boundary Γ' of Δ' contains no poles of $f(z)$, but $I(\Gamma')$ contains the same number of poles as the original parallelogram Δ. This situation is illustrated in Figure 4.4, where there are 5 poles (indicated by heavy dots) in Δ, 8 poles on Γ, 2 poles in $I(\Gamma)$ and 5 poles in $I(\Gamma')$.

---

[10] In stating this definition, we tacitly rely upon the result of Problem 4.15.

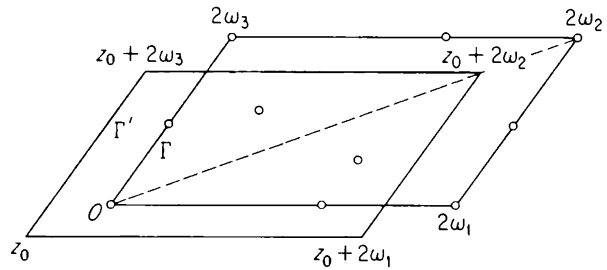


FIGURE 4.4

It follows that

$$\sum_{\Delta} \operatorname{Res} f(z) = \sum_{I(\Gamma')} \operatorname{Res} f(z),$$

where

$$\sum_{E} \operatorname{Res} f(z)$$

denotes the sum of the residues of $f(z)$ at its poles in the set E. But $f(z)$ has no poles on Γ', and hence by Theorem II.2.1,

$$\sum_{\Delta} \operatorname{Res} f(z) = \frac{1}{2\pi i} \int_{\Gamma'} f(z)\, dz. \tag{4.20}$$

Calculating the integral in (4.20), we find that

$$\begin{aligned}\int_{\Gamma'} f(z)\, dz &= \int_{z_0}^{z_0 + 2\omega_1} f(z)\, dz + \int_{z_0 + 2\omega_1}^{z_0 + 2\omega_2} f(z)\, dz \\ &\quad + \int_{z_0 + 2\omega_2}^{z_0 + 2\omega_3} f(z)\, dz + \int_{z_0 + 2\omega_3}^{z_0} f(z)\, dz \\ &= \int_{z_0}^{z_0 + 2\omega_1} f(z)\, dz + \int_{z_0 + 2\omega_1}^{z_0 + 2\omega_2} f(z)\, dz \\ &\quad - \int_{z_0 + 2\omega_3}^{z_0 + 2\omega_2} f(z)\, dz - \int_{z_0}^{z_0 + 2\omega_3} f(z)\, dz,\end{aligned}$$

where

$$\int_a^b f(z)\, dz \equiv \int_{[a,b]} f(z)\, dz.$$

Therefore

$$\begin{aligned}\int_{\Gamma'} f(z)\, dz &= \int_0^1 f(z_0 + 2\omega_1 t)\, dt + \int_0^1 f(z_0 + 2\omega_1 + (2\omega_2 - 2\omega_1)t)\, dt \\ &\quad - \int_0^1 f(z_0 + 2\omega_3 + (2\omega_2 - 2\omega_3)t)\, dt - \int_0^1 f(z_0 + 2\omega_3 t)\, dt \\ &= \int_0^1 f(z_0 + 2\omega_1 t)\, dt + \int_0^1 f(z_0 + 2\omega_1 + 2\omega_3 t)\, dt \\ &\quad - \int_0^1 f(z_0 + 2\omega_3 + 2\omega_1 t)\, dt - \int_0^1 f(z_0 + 2\omega_1 + 2\omega_3 t)\, dt \\ &= \int_0^1 f(z_0 + 2\omega_1 t)\, dt + \int_0^1 f(z_0 + 2\omega_3 t)\, dt \\ &\quad - \int_0^1 f(z_0 + 2\omega_1 t)\, dt - \int_0^1 f(z_0 + 2\omega_3 t)\, dt = 0,\end{aligned}$$

and hence, according to (4.20),

$$\sum_{\Delta} \operatorname{Res} f(z) = 0,$$

as asserted.

THEOREM 4.17. *The order of an elliptic function* $f(z) \not\equiv$ const *is at least* 2.

Proof. If $f(z)$ is of order less than 2, then $f(z)$ has either no poles in Δ or a unique simple pole $\alpha \in \Delta$. In the latter case, the principal part of the Laurent expansion of $f(z)$ at α is of the form

$$\frac{A}{z - \alpha},$$

where A is the residue of $f(z)$ at α. But then $A = 0$ by Theorem 4.16, i.e., $f(z)$ actually has no poles in Δ. Thus, in either case, it follows from Theorem 4.15 that $f(z) \equiv$ const, contrary to hypothesis. This contradiction establishes the theorem.

THEOREM 4.18. *Let* $f(z) \not\equiv$ const *be an elliptic function. Then* $n(A)$, *the total number of A-points of* $f(z)$ *in* Δ, *each counted a number of times equal to its order,*[11] *is independent of* Δ *and equal to the order of* $f(z)$. *In particular,* $n(A) \geqslant 2$.

Proof. If $A = \infty$, the theorem is an immediate consequence of Definition 2, since $n(\infty)$ is just the number of poles in Δ. If $A \neq \infty$, let Γ' be the same as in the proof of Theorem 4.16, except that now ρ is the distance between the set of poles and A-points of $f(z)$ in Δ (for fixed A) and the set $[2\omega_1, 2\omega_2] \cup [2\omega_2, 2\omega_3]$.[12] Then, according to Theorem II.2.2,

$$n(A) - n(\infty) = \frac{1}{2\pi i} \int_{\Gamma'} \frac{f'(z)}{f(z) - A}\, dz. \qquad (4.21)$$

But by Theorem 4.14, the function

$$\varphi(z) = \frac{f'(z)}{f(z) - A}$$

is itself an elliptic function, with the same periods $2\omega_1$ and $2\omega_3$ as $f(z)$. Therefore

$$\frac{1}{2\pi i} \int_{\Gamma'} \frac{f'(z)}{f(z) - A}\, dz = \frac{1}{2\pi i} \int_{\Gamma'} \varphi(z)\, dz = 0, \qquad (4.22)$$

as in Theorem 4.16. Combining (4.21) and (4.22), we find that $n(A) = n(\infty)$, as asserted. The fact that $n(A) \geqslant 2$ now follows from Theorem 4.17.

---

[11] Cf. p. I.371.

[12] Again $\rho = 0$, if Γ, the boundary of Δ, contains no poles or A-points.

THEOREM 4.19. *Let $f(z) \not\equiv \text{const}$ be an elliptic function. Then $s(A)$, the sum of all the A-points of $f(z)$ in Δ, is congruent to $s(\infty)$, the sum of all the poles of $f(z)$ in Δ.*[13]

Proof. If $A = \infty$, the theorem is obvious. If $A \neq \infty$, let Γ' be the same as in the proof of Theorem 4.18. Then, according to Theorem II.2.2,

$$s(A) - s(\infty) = \int_{\Gamma'} z \frac{f'(z)}{f(z) - A}\, dz. \tag{4.23}$$

Calculating the integral in (4.23), we find that

$$\begin{aligned}
\int_{\Gamma'} z \frac{f'(z)}{f(z) - A}\, dz &= \int_{z_0}^{z_0 + 2\omega_1} z \frac{f'(z)}{f(z) - A}\, dz + \int_{z_0 + 2\omega_1}^{z_0 + 2\omega_2} z \frac{f'(z)}{f(z) - A}\, dz \\
&\quad - \int_{z_0 + 2\omega_3}^{z_0 + 2\omega_2} z \frac{f'(z)}{f(z) - A}\, dz - \int_{z_0}^{z_0 + 2\omega_3} z \frac{f'(z)}{f(z) - A}\, dz \\
&= 2\omega_1 \int_0^1 (z_0 + 2\omega_1 t) \frac{f'(z_0 + 2\omega_1 t)}{f(z_0 + 2\omega_1 t) - A}\, dt \\
&\quad + 2\omega_3 \int_0^1 (z_0 + 2\omega_1 + 2\omega_3 t) \frac{f'(z_0 + 2\omega_3 t)}{f(z_0 + 2\omega_3 t) - A}\, dt \\
&\quad - 2\omega_1 \int_0^1 (z_0 + 2\omega_3 + 2\omega_1 t) \frac{f'(z_0 + 2\omega_1 t)}{f(z_0 + 2\omega_1 t) - A}\, dt \\
&\quad - 2\omega_3 \int_0^1 (z_0 + 2\omega_3 t) \frac{f'(z_0 + 2\omega_3 t)}{f(z_0 + 2\omega_3 t) - A}\, dt,
\end{aligned}$$

and hence

$$\begin{aligned}
\int_{\Gamma'} z \frac{f'(z)}{f(z) - A}\, dz &= -2\omega_1 \cdot 2\omega_3 \int_0^1 \frac{f'(z_0 + 2\omega_1 t)}{f(z_0 + 2\omega_1 t) - A}\, dt \\
&\quad + 2\omega_3 \cdot 2\omega_1 \int_0^1 \frac{f'(z_0 + 2\omega_3 t)}{f(z_0 + 2\omega_3 t) - A}\, dt \\
&= -2\omega_3 [\operatorname{Ln} f(z_0 + 2\omega_1 t) - A] \Big|_{t=0}^{t=1} \\
&\quad + 2\omega_1 [\operatorname{Ln} f(z_0 + 2\omega_3 t) - A] \Big|_{t=0}^{t=1} \\
&= -2\omega_3 \operatorname{Ln} \frac{f(z_0 + 2\omega_1) - A}{f(z_0) - A} \\
&\quad + 2\omega_1 \operatorname{Ln} \frac{f(z_0 + 2\omega_3) - A}{f(z_0) - A} \\
&= -2\omega_3 \operatorname{Ln} 1 + 2\omega_1 \operatorname{Ln} 1 \\
&= 2\pi i (2m\omega_1 + 2n\omega_3),
\end{aligned} \tag{4.24}$$

where m and n are integers. Combining (4.23) and (4.24), we obtain

$$s(A) - s(\infty) = 2m\omega_1 + 2n\omega_3,$$

i.e., the complex numbers $s(A)$ and $s(\infty)$ differ by a period of $f(z)$ and hence are congruent in the sense of Definition 1, p. 147.

---

[13] As usual, each A-point and pole is counted a number of times equal to its order.

PROBLEMS

4.1. Let $f(z)$ be meromorphic and periodic, with periods ω_1 and ω_2 such that ω_1/ω_2 is an irrational number. Prove that $f(z) \equiv$ const.

4.2. Given n nonzero complex numbers $\omega_1, \ldots, \omega_n$, the set of all points of the form $m_1\omega_1 + \cdots + m_n\omega_n$, where $m_1, \ldots, m_n$ are arbitrary integers, is called the *lattice* generated by $\omega_1, \ldots, \omega_n$, and the numbers $\omega_1, \ldots, \omega_n$ are called a *basis* of the lattice. Prove that no lattice has a unique basis (cf. Theorem 4.6).

4.3. A lattice generated by a basis B is said to be *reducible* if it is also generated by a basis B' consisting of fewer elements than B. Prove that the lattice generated by $\omega_1, \ldots, \omega_n$ is reducible if and only if there exist n integers $\mu_1, \ldots, \mu_n$, not all zero, such that

$$\mu_1\omega_1 + \cdots + \mu_n\omega_n = 0.$$

Give an example of an irreducible lattice generated by a basis consisting of three elements.

4.4. Let λ be an irreducible lattice generated by a basis consisting of more than two elements. Prove that λ must have a finite limit point.

Comment. Thus there is no meromorphic function $f(z) \not\equiv$ const taking the same value at every point of λ, and we again find that every periodic meromorphic function $f(z) \not\equiv$ const is either simply periodic or doubly periodic (see Theorem 4.5).

4.5. Prove that a periodic entire function $f(z) \not\equiv$ const must be transcendental.

4.6. A function $f(z)$ defined for all z is said to satisfy an *algebraic addition theorem* if, given any z_1 and z_2, the values

$$u = f(z_1), \qquad v = f(z_2), \qquad w = f(z_1 + z_2)$$

satisfy a relation of the form

$$P(u, v, w) = 0,$$

where $P(u, v, w)$ is a polynomial in the three variables u, v and w. Find the polynomial $P(u, v, w)$ if $f(z)$ is given by

a) az; b) az^2; c) e^z; d) $\cos z$; e) $\sin z$.

Ans. b) $(w - u - v)^2 - 4uv$; d) $w^2 - 2uvw + u^2 + v^2 - 1$.

4.7. Prove that every trigonometric polynomial satisfies an algebraic addition theorem.

4.8. Prove that if an entire transcendental function $f(z)$ of finite order satisfies an algebraic addition theorem, then $f(z)$ is periodic.

Hint. Let u, v and w be the same as in Problem 4.6, and suppose the addition theorem is of the form

$$P(u, v, w) = p_n(u, v)w^n + p_{n-1}(u, v)w^{n-1} + \cdots + p_0(u, v) = 0. \qquad (4.25)$$

Let a be such that $p_n(a, v) \not\equiv 0$, and moreover suppose a is not Picard's exceptional value for $f(z)$. Then by Theorem II.9.11, the equation $f(z) = a$ has $n + 1$ distinct roots $c_1, \ldots, c_{n+1}$ (in fact, infinitely many roots). Setting

$$u = f(c_j) = a, \qquad v = f(z), \qquad w = f(c_j + z)$$

in (4.25), we find that

$$p_n[a, f(z)][f(c_j + z)]^n + p_1[a, f(z)][f(c_j + z)]^{n-1} + \cdots + p_0[a, f(z)] = 0$$

for $j = 1, \ldots, n + 1$. But for fixed z, the equation

$$p_n[a, f(z)]w^n + p_1[a, f(z)]w^{n-1} + \cdots + p_0[a, f(z)] = 0$$

has at most n distinct roots (assume that z is not a root of $p_n[a, f(z)] = 0$). Therefore at least two of the numbers $f(c_1 + z), \ldots, f(c_{n+1} + z)$ are equal. Repeat this argument for all the points of some set with a finite limit point, and then invoke the interior uniqueness theorem to prove that $f(c_j + z) \equiv f(c_k + z)$ where $c_j \neq c_k$.

Comment. The restriction to functions of finite order is unnecessary (cf. footnote 11, p. II.268). Moreover, it can actually be shown that $f(z)$ is a trigonometric polynomial.

4.9. Prove that if two entire functions $f(z)$ and $g(z)$ satisfy the identity

$$[f(z)]^2 + [g(z)]^2 = 1,$$

then there is an entire function $h(z)$ such that

$$f(z) = \cos [h(z)], \qquad g(z) = \sin [h(z)].$$

4.10. Prove that there are no nonconstant entire functions $f(z)$ and $g(z)$ satisfying the identity

$$[f(z)]^n + [g(z)]^n = 1$$

if $n \geqslant 3$.

Hint. If $\varepsilon = \cos(\pi/n) + i \sin(\pi/n)$, then

$$[f(z)]^n + [g(z)]^n = [f(z) - \varepsilon g(z)][f(z) - \varepsilon^3 g(z)] \cdots [f(z) - \varepsilon^{2n-1} g(z)],$$

where none of the factors on the right can vanish. Eventually use the result of the preceding problem.

4.11. Let $f(z)$ be meromorphic and periodic with period 2π, and suppose $f(z)$ has only finitely many poles in the period strip g defined on p. 141. Suppose $f(z)$ remains bounded as z approaches the upper end of g. Find series expansions for $f^*(t)$ and $f(z)$, where $f^*(t)$ is the same as in Theorem 4.7.

Hint. If $t = e^{iz} = e^{i(x+iy)}$, then $|t| = e^{-y} \to 0$ as $y \to +\infty$, and hence $t = 0$ is a regular point of $f^*(t)$.

Ans.

$$f^*(t) = \sum_{k=0}^{\infty} a_k t^k \text{ for } |t| < r, \quad f(z) = \sum_{k=0}^{\infty} a_k e^{ikz} \text{ for } y > \ln \frac{1}{r}.$$

4.12. Let $f(z)$, $f^*(t)$ and g be the same as in Problem 4.11, but suppose $f(z)$ remains bounded as z approaches the lower end of g. Find series expansions for $f^*(t)$ and $f(z)$.

Hint. As $y \to -\infty$, $|t| \to \infty$, and hence $t = \infty$ is a regular point of $f^*(t)$.

Ans.

$$f^*(t) = \sum_{k=0}^{\infty} b_k t^{-k} \text{ for } |t| > R, \quad f(z) = \sum_{k=0}^{\infty} b_k e^{-ikz} \text{ for } y < \ln \frac{1}{R}.$$

4.13. With the same notation as in Problems 4.11 and 4.12, find series expansions for $f^*(t)$ and $f(z)$ if

a) $f(z) \to \infty$ as z approaches the upper (or lower) end of g;

b) $f(z)$ approaches no finite or infinite limit as z approaches the upper (or lower) end of g.

Hint. In the first case, $f^*(t)$ has a pole at $t = 0$ (or at $t = \infty$), while in the second case, $f^*(t)$ has an essential singular point at $t = 0$ (or at $t = \infty$).

4.14. Are there elliptic functions of infinite order?

4.15. Prove that an elliptic function has the same number of poles in any fundamental period parallelogram.

4.16. Prove that if $f(z)$ is an odd elliptic function, with period ω, then $\frac{1}{2}\omega$ is either a zero or a pole of $f(z)$, in fact a zero or pole of odd order.

4.17. Let $f(z)$ and $g(z)$ be two elliptic functions with the same fundamental periods. Prove that $f(z)$ and $g(z)$ satisfy an algebraic relation of the form

$$P[f(z), g(z)] \equiv \text{const},$$

where $P(u, v)$ is a polynomial in u and v with constant coefficients.

Hint. If the degree of $P(u, v)$ is high enough (how high?), the coefficients can be chosen in such a way that $P[f(z), g(z)]$ has no poles. But then $P[f(z), g(z)] \equiv \text{const}$, by Theorem 4.15.

5

ELLIPTIC FUNCTIONS: WEIERSTRASS' THEORY

21. Weierstrass' Elliptic Functions

The general properties of elliptic functions were explored in Sec. 20. We now undertake the explicit construction of such functions.

DEFINITION. *Let* ω_1 *and* ω_3 *be two complex numbers such that the points* 0, ω_1 *and* ω_3 *are noncollinear, and let* $f(\Omega)$ *be a function defined for all numbers of the form*

$$\Omega = 2m\omega_1 + 2n\omega_3 \qquad (m, n = 0, \pm 1, \pm 2, \ldots).$$

Then

$$\sum f(\Omega)$$

is shorthand for

$$\sum_{m,n=-\infty}^{\infty} f(2m\omega_1 + 2n\omega_3),$$

while

$$\sum{}' f(\Omega)$$

(with a prime on the summation sign) is shorthand for the same series, with the term corresponding to $m = n = 0$ *(i.e.,* $\Omega = 0$*) omitted.*

LEMMA 1. *The series*

$$\sum{}' \frac{1}{|\Omega|^{\lambda}} \tag{5.1}$$

converges for $\lambda > 2$ *and diverges for* $\lambda \leqslant 2$.

Proof. Let π_k be the boundary of the parallelogram with vertices

$$2k\omega_1 + 2k\omega_3, \quad -2k\omega_1 + 2k\omega_3, \quad -2k\omega_1 - 2k\omega_3, \quad 2k\omega_1 - 2k\omega_3,$$

and let

$$d = \min_{z \in \pi_1} |z|, \qquad D = \max_{z \in \pi_1} |z|.$$

Then clearly

$$\min_{z \in \pi_k} |z| = kd, \qquad \max_{z \in \pi_k} |z| = kD. \tag{5.2}$$

Moreover, π_k contains $8k$ points of the form $\Omega = 2m\omega_1 + 2n\omega_3$. In fact, there are $2k + 1$ such points on each side of π_k, but each of the four vertices of π_k is a point of the form Ω shared by two sides of π_k. Therefore π_k contains a total of

$$4(2k + 1) - 4 = 8k$$

points of the form Ω, as shown in Figure 5.1 for $k = 1$ and 2. Because of (5.2), the contribution of the $8k$ points on π_k to the sum (5.1) lies between $8/D^\lambda k^{\lambda-1}$ and $8/d^\lambda k^{\lambda-1}$. It follows that

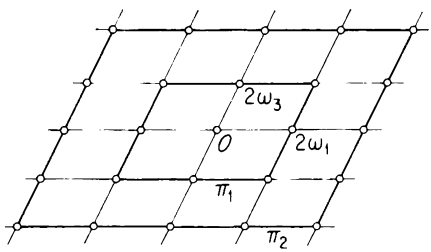


FIGURE 5.1

$$\frac{8}{D^\lambda} \sum_{k=1}^{\infty} \frac{1}{k^{\lambda-1}} \leqslant \sum \frac{1}{|\Omega|^\lambda} \leqslant \frac{8}{d^\lambda} \sum_{k=1}^{\infty} \frac{1}{k^{\lambda-1}}.$$

Since

$$\sum_{k=1}^{\infty} \frac{1}{k^{\lambda-1}}$$

converges for $\lambda > 2$ and diverges for $\lambda \leqslant 2$, the lemma is proved.

COROLLARY. *The exponent of convergence*[1] *of the series* (5.1) *equals* 2.

THEOREM 5.1. *The series*

$$f(z) = \sum \frac{1}{(z - \Omega)^3} \tag{5.3}$$

represents a meromorphic function, whose poles are all of order 3 *and coincide with the set of points* $\Omega = 2m\omega_1 + 2n\omega_3$ $(m, n = 0, \pm 1, \pm 2, \ldots)$.

Proof. Writing

$$f(z) = \sum_{|\Omega| \leqslant 2R} \frac{1}{(z - \Omega)^3} + \sum_{|\Omega| > 2R} \frac{1}{(z - \Omega)^3},$$

[1] See Sec. II.47.

we find that the series

$$\sum_{|\Omega|>2R} \frac{1}{(z-\Omega)^3} \tag{5.4}$$

is absolutely and uniformly convergent on the disk $K_R: |z| < R$, since

$$\sum_{|\Omega|>2R} \frac{1}{(z-\Omega)^3} \leqslant \sum_{|\Omega|>2R} \frac{1}{(|\Omega|-|z|)^3}$$
$$= \sum_{|\Omega|>2R} \frac{1}{\left(1-\frac{|z|}{|\Omega|}\right)^3 |\Omega|^3} < \sum_{|\Omega|>2R} \frac{8}{|\Omega|^3} \qquad (z \in K_R),$$

where the last series on the right converges because of Lemma 1. Therefore (5.4) represents an analytic function on K_R (see Theorem I.15.6). But (5.4) differs from $f(z)$ by the rational function

$$\sum_{|\Omega|\leqslant 2R} \frac{1}{(z-\Omega)^3},$$

whose only singular points are poles of order 3, at the points Ω such that $|\Omega| \leqslant 2R$. Therefore $f(z)$ is meromorphic on K_R, with poles at the points Ω contained in K_R. Since $R > 0$ is arbitrary, $f(z)$ is meromorphic in the whole z-plane, with poles of order 3 at the points Ω, as asserted.

THEOREM 5.2. *The function* (5.3) *is an odd elliptic function of order* 3, *with fundamental periods* $2\omega_1$ *and* $2\omega_3$.

Proof. Writing

$$f(-z) = \sum \frac{1}{(-z-\Omega)^3} = -\sum \frac{1}{[z-(-\Omega)]^3}, \tag{5.5}$$

we observe that the set of all numbers $-\Omega$ coincides with the set of all numbers Ω. In fact, the transformation $\Omega \to -\Omega$, corresponding to reflection in the origin, carries the lattice of points $\Omega = 2m\omega_1 + 2n\omega_3$ into itself. Therefore the two absolutely convergent series

$$\sum \frac{1}{[z-(-\Omega)]^3}, \qquad \sum \frac{1}{(z-\Omega)^3}$$

differ only in the order of their terms, and hence have the same sum $f(z)$. It follows that

$$f(-z) = -f(z),$$

i.e., $f(z)$ is odd, as asserted.

Similarly, writing

$$f(z+2\omega_j) = \sum \frac{1}{[z-(\Omega-2\omega_j)]^3} \qquad (j = 1, 3),$$

we observe that the set of all numbers $\Omega - 2\omega_j$ coincides with the set of all numbers Ω. In fact, the transformation $\Omega \to \Omega - 2\omega_j$, corresponding to displacement of the whole plane by the vector $2\omega_j$, carries the lattice of points $\Omega = 2m\omega_1 + 2n\omega_3$ into itself. Therefore the two absolutely convergent series

$$\sum \frac{1}{[z - (\Omega - 2\omega_j)]^3}, \qquad \sum \frac{1}{(z - \Omega)^3}$$

differ only in the order of their terms, and hence have the same sum $f(z)$. It follows that

$$f(z + 2\omega_j) = f(z) \qquad (j = 1, 3),$$

i.e., $2\omega_1$ and $2\omega_3$ are periods of $f(z)$. In particular, $f(z)$ is elliptic, being doubly periodic and meromorphic. Moreover, if ω is any period of $f(z)$, then, since $\Omega = 2m\omega_1 + 2n\omega_3$ is a pole of $f(z)$, so is $\Omega + \omega$. Therefore

$$\Omega + \omega = \Omega' = 2m'\omega_1 + 2n'\omega_3,$$

which implies

$$\omega = \Omega' - \Omega = 2(m' - m)\omega_1 + 2(n' - n)\omega_3.$$

Thus $2\omega_1$ and $2\omega_3$ are *fundamental* periods of $f(z)$. Of the four points 0, $2\omega_1$, $2\omega_2 = 2(\omega_1 + \omega_3)$, $2\omega_3$ forming a fundamental period parallelogram Δ, only the point 0 is assigned to Δ (recall the convention of Sec. 20), and the other three points are assigned to neighboring parallelograms. Therefore the point 0 is the only pole in Δ. Since this pole is of order 3, $f(z)$ is an elliptic function of order 3, as asserted.

Next we use $f(z)$ to construct an elliptic function of order 2. Let z_0 be any point of the z-plane, which is not a pole of $f(z)$. Then, integrating (5.3) term by term along any rectifiable curve γ which joins z_0 to another point z and passes through no poles of $f(z)$, we obtain a new function

$$\varphi(z) = C + \int_{z_0}^{z} f(z)\,dz = C - \frac{1}{2}\sum \left[\frac{1}{(z - \Omega)^2} - \frac{1}{(z_0 - \Omega)^2}\right]. \tag{5.6}$$

Since the series in (5.3) is uniformly convergent on every compact set containing no points of the form $\Omega = 2m\omega_1 + 2n\omega_3$, the same is true of the series in (5.6).[2] Therefore $\varphi(z)$ is meromorphic with a pole at every point Ω. Writing $\varphi(z)$ in the form

$$\varphi(z) = C - \frac{1}{2z^2} + \frac{1}{2z_0^2} - \frac{1}{2}\sum{}' \left[\frac{1}{(z - \Omega)^2} - \frac{1}{(z_0 - \Omega)^2}\right], \tag{5.7}$$

[2] See Problem I.17.7.

where, as usual, the prime on the summation sign indicates omission of the term corresponding to $\Omega = 0$, we see that

$$\varphi(z) + \frac{1}{2z^2}$$

is a meromorphic function which is regular at the origin, where it takes the value

$$\left[\varphi(z) + \frac{1}{2z^2}\right]\Big|_{z=0} = C + \frac{1}{2z_0^2} - \frac{1}{2}\sum{}' \left[\frac{1}{\Omega^2} - \frac{1}{(z_0 - \Omega)^2}\right]. \tag{5.8}$$

Choosing the constant C to make (5.8) vanish, and then subtracting (5.8) from (5.7), we obtain

$$\varphi(z) = -\frac{1}{2}\left\{\frac{1}{z^2} + \sum{}' \left[\frac{1}{(z - \Omega)^2} - \frac{1}{\Omega^2}\right]\right\}.$$

The function appearing in braces [equal to $-2\varphi(z)$] was introduced by Weierstrass, and is denoted by $\wp(z)$. Thus, by definition,

$$\wp(z) = \frac{1}{z^2} + \sum{}' \left[\frac{1}{(z - \Omega)^2} - \frac{1}{\Omega^2}\right]. \tag{5.9}$$

Clearly $\wp(z)$ is a meromorphic function, whose poles are all of order 2 and coincide with the set of points $\Omega = 2m\omega_1 + 2n\omega_3$ ($m, n = 0, \pm 1, \pm 2, \ldots$). At each point Ω, the principal part of $\wp(z)$ is given by

$$\frac{1}{(z - \Omega)^2}.$$

The function $\wp(z)$ and its derivative $\wp'(z)$ play a key role in the theory of elliptic functions, and are called *Weierstrass' elliptic functions.*

Remark. By the very construction of $\wp(z)$ [cf. (5.9)], the function

$$\wp(z) - \frac{1}{z^2}$$

vanishes at the origin.

LEMMA 2. *The series* (5.9) *is absolutely convergent.*

Proof. Omitting a finite number of terms has no effect on the absolute convergence of (5.9). Considering only terms such that $|\Omega| > 2|z|$, we have

$$\left|\frac{1}{(z - \Omega)^2} - \frac{1}{\Omega^2}\right| = \left|\frac{(2\Omega - z)z}{\Omega^2(z - \Omega)^2}\right| \leqslant \frac{2|\Omega|\left(1 + \dfrac{|z|}{2|\Omega|}\right)|z|}{|\Omega|^4\left(1 - \dfrac{|z|}{|\Omega|}\right)^2} < \frac{10|z|}{|\Omega|^3}.$$

The proof now follows from Lemma 1 and the comparison test.

THEOREM 5.5. *The function $\wp(z)$ is an even elliptic function of order* 2, *with fundamental periods* $2\omega_1$ *and* $2\omega_3$.

Proof. Writing

$$\wp(-z) = \frac{1}{z^2} + \sum{}' \left[\frac{1}{(z+\Omega)^2} - \frac{1}{\Omega^2}\right]$$
$$= \frac{1}{z^2} + \sum{}' \left[\frac{1}{\{z-(-\Omega)\}^2} - \frac{1}{(-\Omega)^2}\right],$$

we observe that the second series on the right and the series in (5.9) differ only in the order of their terms, and hence have the same sum $\wp(z)$, being absolutely convergent (see Lemma 2). It follows that

$$\wp(-z) = \wp(z),$$

i.e., $\wp(z)$ is even, as asserted. The derivative $\wp'(z)$ is given by

$$\wp'(z) = -\frac{2}{z^3} - \sum{}' \frac{2}{(z-\Omega)^3} = -2\sum \frac{1}{(z-\Omega)^3} = -2f(z),$$

where $f(z)$ is the function (5.3). According to Theorem 5.2, $f(z)$ is doubly periodic with periods $2\omega_1$ and $2\omega_3$. Therefore

$$\wp'(z+2\omega_j) - \wp'(z) = 0 \qquad (j = 1, 3),$$

or, after integration,

$$\wp(z+2\omega_j) - \wp(z) = C_j,$$

and hence

$$C_j = 0 \qquad (j = 1, 3),$$

since $\wp(z)$ is even. It follows that

$$\wp(z+2\omega_j) = \wp(z),$$

i.e., $\wp(z)$ is doubly periodic, with periods $2\omega_1$ and $2\omega_3$ which are fundamental periods by the same argument as for $f(z)$. Of the four points 0, $2\omega_1$, $2\omega_2$, $2\omega_3$ forming a fundamental period parallelogram, only the point 0 is assigned to Δ. Therefore the point 0 is the only pole in Δ (why?). Since this pole is of order 2, $\wp(z)$ is an elliptic function of order 2, as asserted.

COROLLARY. *The derivative $\wp'(z)$ is an odd elliptic function of order* 3, *with fundamental periods* $2\omega_1$ *and* $2\omega_3$.

Since $\wp'(z)$ is of order 3, it follows from Theorem 4.18 that, given any A, $\wp'(z)$ has three A-points in the fundamental period parallelogram Δ (with vertices 0, $2\omega_1$, $2\omega_2$, $2\omega_3$). If $A = \infty$, these points coalesce into just one

point, i.e., the triple pole of $\wp'(z)$ at the origin. Choosing $A = 0$, we see that $\wp'(z)$ must have three zeros in Δ. These zeros are easily found. In fact,

$$\wp'(z) = -\wp'(z)$$

implies

$$\wp'(2\omega_j - z) = -\wp'(z) \qquad (j = 1, 2, 3),$$

and hence, after setting $z = \omega_j$,

$$\wp'(\omega_j) = -\wp'(\omega_j).$$

But $\wp'(\omega_j) \neq \infty$, and therefore

$$\wp'(\omega_j) = 0 \qquad (j = 1, 2, 3)$$

(cf. Problem 4.16). Thus $\wp'(z)$ has zeros at each of the points ω_1, ω_2 and ω_3. The point ω_2 lies at the center of Δ, while the other two points ω_1 and ω_3 lie at the midpoints of the two sides of $\overline{\Delta}$ assigned to Δ (recall p. 147). Obviously, each of the zeros ω_1, ω_2, ω_3 is simple, since otherwise the total number of zeros of $\wp'(z)$ in Δ would exceed 3, contrary to Theorem 4.18. Note that the sum of the zeros $\omega_1 + \omega_2 + \omega_3 = 2\omega_1 + 2\omega_3$ differs by a period from the sum of the poles $0 + 0 + 0 = 0$, in keeping with Theorem 4.19.

Similarly, since $\wp(z)$ is of order 2, it follows from Theorem 4.18 that, given any A, $\wp(z)$ has two A-points in Δ. If $A = \infty$, these points coalesce into just one point, i.e., the double pole of $\wp(z)$ at the origin. The A-points corresponding to the values $\wp(\omega_j) = e_j$ $(j = 1, 2, 3)$ are also of order 2, since

$$\wp'(\omega_j) = 0, \quad \wp''(\omega_j) \neq 0 \qquad (j = 1, 2, 3).$$

However, for all other values of A ($A \neq \infty$, $A = e_1, e_2, e_3$), the A-points are simple, since otherwise $\wp'(z)$ would have zeros in Δ other than the three points ω_1, ω_2, ω_3, and this is impossible.

To examine the simple A-points in more detail, we note that

$$\wp(-z) = \wp(z)$$

implies

$$\wp(2\omega_j - z) = \wp(z) \qquad (j = 1, 2, 3). \tag{5.10}$$

If $z_0 \in \Delta$ is a simple A-point, then $z_0 \neq 0$, $z_0 \neq \omega_1, \omega_2, \omega_3$. Suppose first that z_0 is an interior point of Δ. Then the point $2\omega_2 - z_0$, which is symmetric to z_0 with respect to ω_2 (the center of Δ), is also an interior point of Δ. Moreover, according to (5.10), $\wp(z)$ takes the same value at the points z_0 and $2\omega_2 - z_0$. On the other hand, suppose z_0 is a boundary point of Δ. Then z_0 belongs to either $[0, 2\omega_1)$ or $[0, 2\omega_3)$, the two sides of Δ meeting at the origin (recall footnote 9, p. 147). In the first case, the point $2\omega_1 - z_0$ also belongs to $[0, 2\omega_1)$ and $\wp(z)$ takes the same value at z_0 and $2\omega_1 - z_0$, while

in the second case, $2\omega_3 - z_0 \in [0, 2\omega_3)$ and $\wp(z)$ takes the same values at z_0 and $2\omega_3 - z_0$.

We summarize these results in the form of a theorem:

THEOREM 5.4. *The function $\wp(z)$ takes the same value at any two interior points of Δ which are symmetric with respect to ω_2, the center of Δ, and at any two boundary points of Δ which are symmetric with respect to ω_1 or ω_3, the midpoints of the two sides of $\overline{\Delta}$ assigned to Δ. There are only four points in Δ which are not symmetric to any other point in Δ, i.e., the points 0, ω_1, ω_2 and ω_3.*[3] *Each of these points is a double A-point (the first a double pole).*

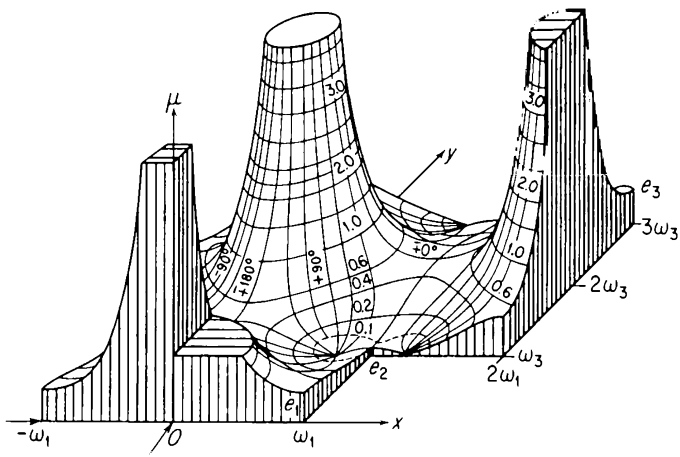


FIGURE 5.2

Figure 5.2 shows the modular surface of the function $\wp(z)$, i.e., the surface $\mu = |\wp(z)|$. The curves lying in horizontal planes belong to the family $\mu = \text{const}$, while the curves labelled in degrees belong to the family $\arg \wp(z) = \text{const}$.

22. The Functions $\wp(z \mid \alpha, i\beta)$ and $\wp(z \mid \alpha - i\beta, \alpha + i\beta)$

So far, ω_1 and ω_3 have been arbitrary complex numbers such that the points 0, ω_1 and ω_3 are noncollinear. We can emphasize the dependence of $\wp(z)$ on ω_1 and ω_3 by writing $\wp(z \mid \omega_1, \omega_3)$ instead of $\wp(z)$. We now study some important special choices of ω_1 and ω_3, using the symbols α and β to denote arbitrary *positive* numbers.

---

[3] Note that the points $2\omega_1$, $2\omega_1 + \omega_3$, $2\omega_2$, $\omega_1 + 2\omega_3$ and $2\omega_3$ do not belong to Δ.

LEMMA 1. *If* $\wp(z) = \wp(z \mid \alpha, i\beta)$, *then*

$$\overline{\wp(x + iy)} = \wp(x - iy). \tag{5.11}$$

Proof. Writing

$$\wp(x + iy) = \frac{1}{(x + iy)^2} + \sum{}' \left[\frac{1}{(x + iy - 2m\alpha - 2ni\beta)^2} - \frac{1}{(2m\alpha + 2ni\beta)^2}\right], \tag{5.12}$$

we find that

$$\overline{\wp(x + iy)} = \frac{1}{(x - iy)^2} + \sum{}' \left[\frac{1}{(x - iy - 2m\alpha + 2ni\beta)^2} - \frac{1}{(2m\alpha - 2ni\beta)^2}\right],$$

while on the other hand,

$$\wp(x - iy) = \frac{1}{(x - iy)^2} + \sum{}' \left[\frac{1}{(x - iy - 2m\alpha - 2ni\beta)^2} - \frac{1}{(2m\alpha + 2ni\beta)^2}\right].$$

Since the set of all numbers of the form $2m\alpha - 2ni\beta$ coincides with the set of all numbers of the form $2m\alpha + 2ni\beta$, the last two series have the same sum.[4] But this implies (5.11), as required.

LEMMA 2. *If* $\wp(z) = \wp(z \mid \alpha, i\beta)$, *then* $\wp(z)$ *takes only real values on the vertical lines* $x = 0$, $x = \alpha$ *and on the horizontal lines* $y = 0$, $y = \beta$.

Proof. We use Lemma 1 repeatedly, together with the evenness and periodicity of $\wp(z)$. Thus, if $x = 0$, then

$$\overline{\wp(iy)} = \wp(-iy)$$

by Lemma 1, while

$$\wp(iy) = \wp(-iy)$$

by evenness, and hence

$$\overline{\wp(iy)} = \wp(iy),$$

i.e., $\wp(iy)$ is real. If $x = \alpha$, then

$$\overline{\wp(\alpha + iy)} = \wp(\alpha - iy),$$

by Lemma 1, while

$$\wp(\alpha + iy) = \wp(-\alpha - iy) = \wp(\alpha - iy),$$

by evenness and periodicity, and hence

$$\overline{\wp(\alpha + iy)} = \wp(\alpha + iy).$$

[4] Again we use the fact that order of terms is irrelevant in the series defining $\wp(z)$, since the series is absolutely convergent (see Lemma 2, p. 159).

If $y = 0$, Lemma 1 gives

$$\overline{\wp(x)} = \wp(x)$$

directly. Finally, if $y = \beta$, we compare

$$\overline{\wp(x + i\beta)} = \wp(x - i\beta)$$

with

$$\wp(x + i\beta) = \wp(x - i\beta).$$

THEOREM 5.5. *If* $\wp(z) = \wp(z \mid \alpha, i\beta)$, *then* $w = \wp(z)$ *maps the rectangle* $G: 0 < x < \alpha$, $0 < y < \beta$ *conformally onto the lower half-plane.*

Proof. According to Theorems 5.3 and 5.4, $\wp(z)$ is one-to-one and wide-sense continuous on $\overline{G}$, and analytic on $\overline{G}$ except at the point $z = 0$. In particular, $w = \wp(z)$ is univalent on G, which it maps conformally onto a domain D in the w-plane. Let γ be the boundary of G. Then the boundary Γ of D is a closed unbounded Jordan curve, and in fact $\Gamma = \wp(\gamma)$. By Lemma 2, Γ is a subset of the real axis. But then Γ must coincide with the whole real axis, if it is to be a closed curve. Therefore D is either the upper or the lower half-plane.

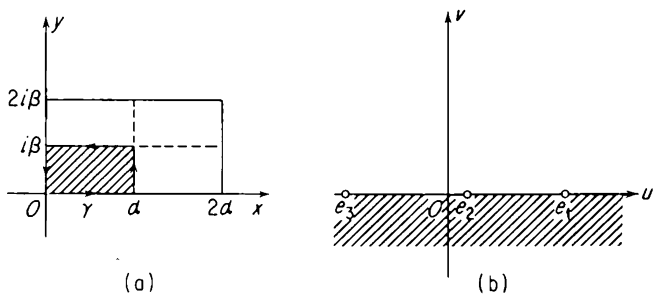


FIGURE 5.3

By a familiar argument (cf. Remark 2, p. I.156), D is the domain on the left of an observer moving with the point $w = \wp(z)$ as z traverses γ once in the positive (i.e., counterclockwise) direction, starting from the point $z = 0$, say [see Figure 5.3(a)]. But when $z = x$ and x is small, the term $1/x^2$ dominates in the expression (5.9) for $\wp(z)$, and when $z = iy$ and y is small, the term $-1/y^2$ dominates. Therefore $w = \wp(z)$ goes from $+\infty$ to $-\infty$ along the real axis as z traverses γ once in the positive direction, and hence the domain D on the left of the moving point $w = \wp(z)$ is the *lower* half-plane, as asserted.

As the point $w = \wp(z)$ moves along the real axis from $+\infty$ to $-\infty$, it successively encounters the points

$$\wp(\omega_1) = \wp(\alpha) = e_1, \quad \wp(\omega_2) = \wp(\alpha + i\beta) = e_2, \quad \wp(\omega_3) = \wp(i\beta) = e_3,$$

and hence $e_1 > e_2 > e_3$ [see Figure 5.3(b)]. Theorem 5.4 and the fact that $\wp(z)$ is an elliptic function of order 2 now immediately imply

THEOREM 5.6. *$\wp(z) = \wp(z \mid \alpha, i\beta)$ takes every value in the interval $[e_1, +\infty]$ twice (with due regard for order) on the (half-open) segment $[0, 2\alpha)$ and every value in the interval $[e_2, e_1]$ twice on the segment $[\alpha, \alpha + 2i\beta)$. Moreover, $\wp(z)$ takes every value in $[e_3, e_2]$ twice on*

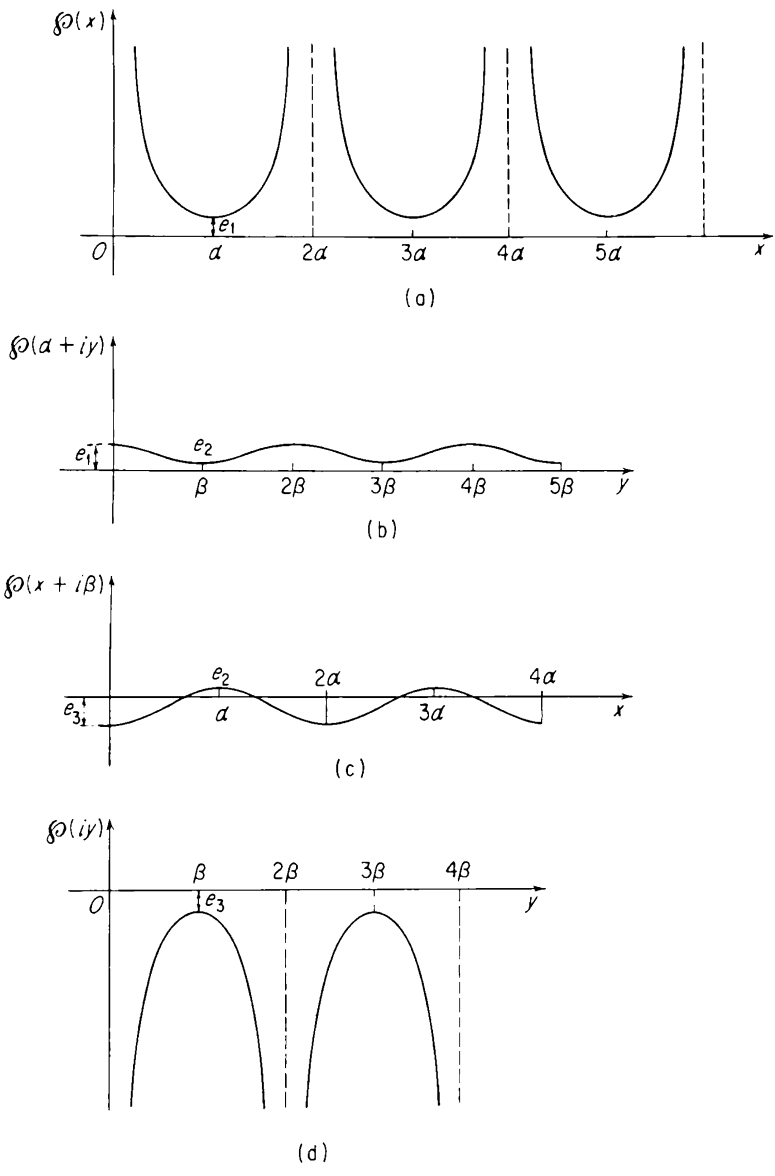


FIGURE 5.4

$[i\beta, 2\alpha + i\beta)$ and every value in $[-\infty, e_3]$ twice on $[0, 2i\beta)$. At every other point of the fundamental period parallelogram Δ,[5] $\wp(z)$ takes imaginary values.

In Figure 5.4 we plot the functions $\wp(x)$, $\wp(\alpha + iy)$, $\wp(x + i\beta)$ and $\wp(iy)$, which together give a complete picture of all real values of $\wp(z) = \wp(z \mid \alpha, i\beta)$. Being periodic, $\wp(z)$ takes the same values on infinitely many other lines parallel to the real and imaginary axes.

Next we consider the case where the fundamental periods $2\omega_1$ and $2\omega_3$ are a pair of conjugate complex numbers:

$$2\omega_1 = 2\alpha - 2i\beta, \quad 2\omega_3 = 2\alpha + 2i\beta \qquad (\alpha, \beta > 0).$$

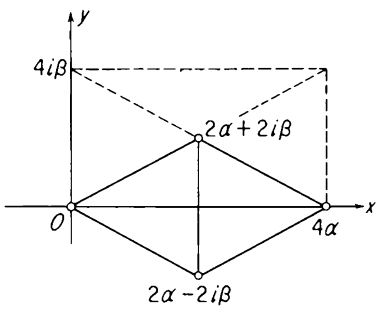


FIGURE 5.5

The fundamental period parallelogram Δ is now a rhombus, with diagonals parallel to the real and imaginary axes. The corresponding function $\wp(z \mid \alpha - i\beta, \alpha + i\beta)$ also has a real period $4\alpha = 2\omega_1 + 2\omega_3$ and a purely imaginary period $4i\beta = -2\omega_1 + 2\omega_3$, but unlike the case of the function $\wp(z \mid \alpha, i\beta)$, these periods are not fundamental, since the rectangle with vertices 0, 4α, $4\alpha + 4i\beta$, $4i\beta$ contains another period at its center, i.e., the period $2\alpha + 2i\beta$ (see Figure 5.5).

LEMMA 1′. *If $\wp(z) = \wp(z \mid \alpha - i\beta, \alpha + i\beta)$, then*

$$\overline{\wp(x + iy)} = \wp(x - iy).$$

Proof. Follow the proof of Lemma 1, starting from the formula

$$\wp(x + iy) = \frac{1}{(x + iy)^2} + \sum{}' \left\{ \frac{1}{[x + iy - 2m(\alpha - i\beta) - 2n(\alpha + i\beta)]^2} - \frac{1}{[2m(\alpha - i\beta) + 2n(\alpha + i\beta)]^2} \right\}, \tag{5.13}$$

instead of (5.12).

LEMMA 2′. *If $\wp(z) = \wp(z \mid \alpha - i\beta, \alpha + i\beta)$, then $\wp(z)$ takes only real values on the lines $x = 2\alpha$ and $y = 0$.*

Proof. If $x = 2\alpha$, then

$$\overline{\wp(2\alpha + iy)} = \wp(2\alpha - iy)$$

by Lemma 1′, while

$$\wp(2\alpha + iy) = \wp(-2\alpha - iy) = \wp(2\alpha - iy)$$

[5] Of course, in this case Δ is a rectangle. Note that the segments $[0, 2\alpha)$, $[\alpha, \alpha + 2i\beta)$, $[i\beta, 2\alpha + i\beta)$ and $[0, 2i\beta)$ are the lower side of Δ, the vertical bisector of Δ, the horizontal bisector of Δ and the left-hand side of Δ, respectively.

by evenness and periodicity. Therefore

$$\overline{\wp(2\alpha + iy)} = \wp(2\alpha + iy),$$

i.e., $\wp(2\alpha + iy)$ is real. If $y = 0$, Lemma 1′ implies

$$\overline{\wp(x)} = \wp(x)$$

directly.

Remark. $\wp(z)$ takes only real values on the imaginary axis, since

$$\wp(iy) = \wp[2\alpha + i(y + 2\beta)].$$

Using Lemma 2′ and Theorem 5.4, we find that as the point z moves along the real axis from 0 to 2α, $\wp(z)$ decreases from $+\infty$ to the value

$$\wp(\omega_2) = \wp(2\alpha) = e_2,$$

since when $z = x$ and x is small, the term $1/x^2$ dominates in the expansion (5.13). Then $\wp(z)$ decreases from e_2 to $-\infty$ as z moves along the segment $x = 2\alpha$, $0 \leqslant y \leqslant 2\beta$, since when $z = 2\alpha + 2i\beta t$ and $t \to 1$, the term

$$\frac{1}{(2\alpha + 2i\beta t - 2\alpha - 2i\beta)^2} = \frac{1}{-4\beta^2(t - 1)^2}$$

dominates in (5.13).

THEOREM 5.6′. *$\wp(z) = \wp(z \mid \alpha - i\beta, \alpha + i\beta)$ takes every value in the interval $[e_2, +\infty]$ twice on the segment $[0, 4\alpha)$ and every value in the interval $[-\infty, e_2]$ twice on the segment $[2\alpha - 2i\beta, 2\alpha + 2i\beta)$. At every other point of the fundamental period parallelogram Δ,[6] $\wp(z)$ takes imaginary values.*

Proof. Use Theorem 5.4 and the fact that $\wp(z)$ is an elliptic function of order 2.

Remark. In particular, the numbers

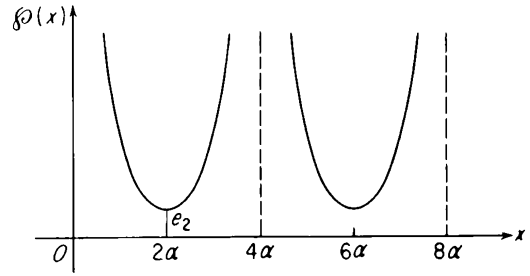
$$e_1 = \wp(\omega_1) = \wp(\alpha - i\beta), \qquad e_3 = \wp(\omega_3) = \wp(\alpha + i\beta)$$

are imaginary and in fact complex conjugates (by Lemma 1′).

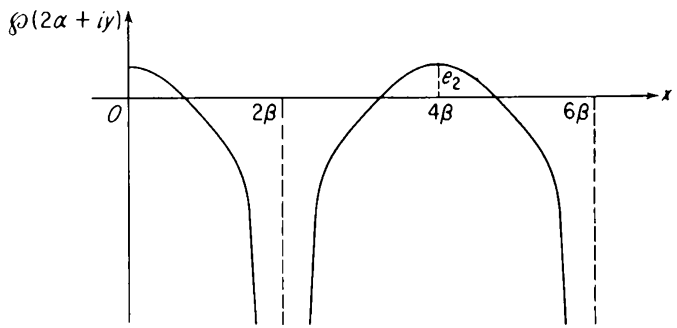
In Figure 5.6 we plot the functions $\wp(x)$ and $\wp(2\alpha + iy)$, which together give a complete picture of all real values of $\wp(z) = \wp(z \mid \alpha - i\beta, \alpha + i\beta)$.

---

[6] Note that the segments $[0, 4\alpha)$ and $[2\alpha - 2i\beta, 2\alpha + 2i\beta)$ are the diagonals of Δ, which is now a rhombus.



(a)



(b)

FIGURE 5.6

23. The Differential Equation for $\wp(z)$

THEOREM 5.7. *The Laurent expansion of* $\wp(z)$ *at the point* $z = 0$ *is*

$$\wp(z) = \frac{1}{z^2} + c_2 z^2 + c_4 z^4 + \cdots + c_{2m} z^{2m} + \cdots, \qquad (5.14)$$

where

$$c_{2m} = (2m + 1) \sum{}' \frac{1}{\Omega^{2m+2}}. \qquad (5.15)$$

Proof. If $\Omega \neq 0$ and $|z| < |\Omega|$, then

$$\frac{1}{(z - \Omega)^2} - \frac{1}{\Omega^2} = \frac{1}{\Omega^2}\left[\left(1 - \frac{z}{\Omega}\right)^{-2} - 1\right] = \sum_{n=1}^{\infty} \frac{n + 1}{\Omega^{n+2}} z^n. \qquad (5.16)$$

Substitution of (5.16) into (5.9) gives[7]

$$\wp(z) = \frac{1}{z^2} + \sum{}' \left[\frac{2z}{\Omega^3} + \frac{3z^2}{\Omega^4} + \cdots + \frac{(n + 1)z^n}{\Omega^{n+2}} + \cdots\right]$$

$$= \frac{1}{z^2} + 2\sum{}' \frac{1}{\Omega^3} z + 3\sum{}' \frac{1}{\Omega^4} z^2 + \cdots + (n + 1)\sum{}' \frac{1}{\Omega^{n+2}} z^n + \cdots$$

[7] This is justified by Theorem I.18.1.

if $0 < |z| < \delta$, where δ is the shortest distance from the origin to a non-zero period Ω. Since $\wp(z)$ is even, all the coefficients c_{2m+1} vanish, and we obtain (5.14) and (5.15).

COROLLARY. *The Laurent expansion of $\wp'(z)$ at the point $z = 0$ is*

$$\wp'(z) = -\frac{2}{z^3} + 2c_2 z + 4c_4 z^3 + \cdots + 2mc_{2m} z^{2m-1} + \cdots. \tag{5.17}$$

Proof. Differentiate (5.14) term by term.

THEOREM 5.8. *The function $\wp(z)$ satisfies the first-order differential equation*

$$[\wp'(z)]^2 = 4\wp^3(z) - g_2\wp(z) - g_3, \tag{5.18}$$

in terms of the quantities

$$g_2 = 60 \sum{}' \frac{1}{\Omega^4}, \qquad g_3 = 140 \sum{}' \frac{1}{\Omega^6}, \tag{5.19}$$

called the invariants of $\wp(z)$.[8]

Proof. Using Theorem 5.7 and its corollary, we find that[9]

$$\begin{aligned}
&[\wp'(z)]^2 - 4\wp^3(z) \\
&\quad = \left(\frac{4}{z^6} - \frac{8c_2}{z^2} - 16c_4 + \cdots\right)^2 - 4\left(\frac{1}{z^6} + \frac{3c_2}{z^2} + 3c_4 + \cdots\right)^3 \\
&\quad = -\frac{20c_2}{z^2} - 28c_4 - \cdots,
\end{aligned}$$

where the unwritten terms all involve positive powers of z, and hence

$$\varphi(z) \equiv [\wp'(z)]^2 - 4\wp^3(z) + 20c_2\wp(z) = -28c_4 + \cdots.$$

But $\varphi(z)$ is an elliptic function (see Theorem 4.14) with no poles in the fundamental period parallelogram. Therefore, by Theorem 4.15,

$$\varphi(z) \equiv \text{const},$$

and in fact,

$$\varphi(z) \equiv \varphi(0) = -28c_4.$$

This is equivalent to (5.18) if

$$g_2 = 20c_2, \qquad g_3 = 28c_4. \tag{5.20}$$

Comparing (5.15) and (5.20), we obtain (5.19).

---

[8] Concerning the meaning of the term "invariants" in this context, see pp. 31, 68–69 of the book by Tricomi (cited in the Bibliography on pp. 351–353).

[9] To justify the formal operations, use Theorem I.3.7 and the absolute convergence of the Laurent series.

Setting $w = \wp(z)$, we can write (5.18) in the form

$$\left(\frac{dw}{dz}\right)^2 = 4w^3 - g_2w - g_3. \tag{5.21}$$

As shown on p. 161, the function $\wp'(z)$ vanishes at the points ω_1, ω_2 and ω_3. Therefore $4w^3 - g_2w - g_3$ vanishes at

$$w_j = \wp(\omega_j) = e_j \qquad (j = 1, 2, 3),$$

and hence

$$4w^3 - g_2w - g_3 = 4(w - e_1)(w - e_2)(w - e_3). \tag{5.22}$$

Since the numbers e_1, e_2 and e_3 are distinct, the discriminant D of the cubic equation

$$4w^3 - g_2w - g_3 = 0$$

must be nonzero, i.e.,[10]

$$D = \tfrac{1}{16}(g_2^3 - 27g_3^2) = (e_2 - e_3)^2(e_3 - e_1)^2(e_1 - e_2)^2 \neq 0. \tag{5.23}$$

It follows from (5.22) that

$$\begin{aligned} e_1 + e_2 + e_3 &= 0, \\ e_1e_2 + e_2e_3 + e_3e_1 &= -\frac{g_2}{4}, \\ e_1e_2e_3 &= \frac{g_3}{4}. \end{aligned} \tag{5.24}$$

For a given value of w, (5.21) leads to values of dw/dz differing only in sign. This is to be expected, since $\wp'(z)$ is odd while $\wp(z)$ is even. However, $\wp'(z)$ is a single-valued function of z, and hence in the formula

$$\wp'(z) = \sqrt{4\wp^3(z) - g_2\wp(z) - g_3}, \tag{5.25}$$

implied by (5.18), one must always be careful to choose the sign leading to the given value of $\wp'(z)$.[11]

Now let γ be any rectifiable curve which joins a fixed point z_0 to a variable point z and passes through no poles of $\wp(z)$. Writing (5.25) as

$$dz = \frac{d\wp(z)}{\sqrt{4\wp^3(z) - g_2\wp(z) - g_3}}$$

and integrating along γ, we obtain

$$z - z_0 = \int_\gamma \frac{d\wp(z)}{\sqrt{4\wp^3(z) - g_2\wp(z) - g_3}}. \tag{5.26}$$

---

[10] See e.g., G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, third edition, The Macmillan Co., New York (1965), p. 106, where it is also shown that the cubic equation has three real roots if $D > 0$ and one real root and two complex conjugate roots if $D < 0$.

[11] The multiple-valued function in the right-hand side of (5.25) has already been encountered on p. I.223.

Let $\Gamma = \wp(\gamma)$ be the image of γ in the w-plane, with end points $w_0 = \wp(z_0)$ and $w = \wp(z)$. Then (5.26) can also be written in the form

$$z - z_0 = \int_{w_0}^{w} \frac{dt}{\sqrt{4t^3 - g_2 t - g_3}}, \tag{5.27}$$

where the integral is along the curve Γ and defines a multiple-valued function of w. In fact e_1, e_2 and e_3 are singular points of the integrand in (5.27), being zeros of the polynomial $4t^3 - g_2 t - g_3$, and hence in general $I \neq I'$, where

$$I = \int_{\Gamma} \frac{dt}{\sqrt{4t^3 - g_2 t - g_3}}, \qquad I' = \int_{\Gamma'} \frac{dt}{\sqrt{4t^3 - g_2 t - g_3}}$$

and Γ, Γ' are distinct paths joining w_0 to w.[12] This is in keeping with the fact that infinitely many distinct values of z correspond to the same value $w = \wp(z)$. These distinct values are all given by the formula

$$z = \pm z' + 2m\omega_1 + 2n\omega_3 \qquad (m, n = 0, \pm 1, \pm 2, \ldots),$$

where z' is an arbitrary w-point of $\wp(z)$, and joining z_0 to any of these points z by a rectifiable curve γ, we find that

$$z - z_0 = \int_{\Gamma} \frac{dt}{\sqrt{4t^3 - g_2 t - g_3}},$$

where $\Gamma = \wp(\gamma)$.

If $z_0 \to 0$ in (5.27), then $w_0 = \wp(z_0) \to \infty$, and hence

$$z = \int_{\infty}^{w} \frac{dt}{\sqrt{4t^3 - g_2 t - g_3}}, \tag{5.28}$$

since the improper integral on the right converges. Here $w = \wp(z)$ and the square root is chosen to agree with $\wp'(z)$. The right-hand side of (5.28) is called an *elliptic integral of the first kind (in Weierstrass' form)*, and the function $w = \wp(z)$ is clearly the inverse of this integral.

Remark 1. More generally, an *elliptic integral* is any expression of the form

$$\int_{w_0}^{w} R(t, \sqrt{a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4})\, dt, \tag{5.29}$$

where $R(t, \tau)$ is a rational function, and the polynomial under the radical is of the fourth degree ($a_4 \neq 0$) or of the third degree ($a_4 = 0$, $a_3 \neq 0$). Such expressions come up in problems involving ellipses (hence the word "elliptic"). For example, let l be the arc length of the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

[12] However, $I = I'$ if there is a simply connected domain G such that $\Gamma \subset G$, $\Gamma' \subset G$ and $e_1, e_2, e_3 \notin G$ (describe other circumstances under which $I = I'$).

measured from the point $(0, b)$ to a variable point (x, y), $y > 0$. Then

$$l = \int_0^x \sqrt{1 + y'^2}\, dx = \int_0^x \sqrt{\frac{a^2 - (a^2 - b^2)\, x^2/a^2}{a^2 - x^2}}\, dx = a \int_a^{x/a} \sqrt{\frac{1 - k^2t^2}{1 - t^2}}\, dt$$

$$= a \int_0^{x/a} \frac{1 - k^2t^2}{\sqrt{(1 - t^2)(1 - k^2t^2)}}\, dt,$$

where

$$k^2 = \frac{a^2 - b^2}{b^2},$$

and the last integral on the right is of the form (5.29). By the same token, functions obtained by inversion of elliptic integrals have also come to be known as "elliptic."

Remark 2. Setting $z = \omega_j$ in (5.28) and noting that $w = \wp(\omega_j) = e_j$, we find that

$$\omega_j = \int_\infty^{e_j} \frac{dt}{\sqrt{4t^3 - g_2t - g_3}} \qquad (j = 1, 2, 3),$$

where the path of integration is the image under $w = \wp(z)$ of any rectifiable curve joining the point $z = 0$ to the point $z = \omega_j$. By varying this path we generate all possible "half-periods" of the form

$$\omega_j + 2m\omega_1 + 2n\omega_3 \qquad (j = 1, 2, 3),$$

where m and n are arbitrary integers. The sign associated with the square root does not matter, since it only affects the sign of the half-period.

24. Inversion of Elliptic Integrals

The considerations of the preceding section naturally raise the following question: Does there always exist a Weierstrass elliptic function $w = \wp(z)$ satisfying the differential equation

$$\left(\frac{dw}{dz}\right)^2 = 4w^3 - g''w - g''', \tag{5.30}$$

where g'' and g''' are given complex numbers? Clearly g'' and g''' must satisfy the condition

$$g''^3 - 27g'''^2 \neq 0,$$

obeyed by the invariants of a Weierstrass function. However, is this condition sufficient?

The problem just posed is equivalent to *inversion* of the elliptic integral

$$z = \int_{\infty}^{w} \frac{dt}{\sqrt{4t^3 - g''t - g'''}}, \tag{5.31}$$

i.e., to showing that the inverse of (5.31) is an elliptic function $w = \wp(z)$ with invariants g'' and g'''. To solve the problem, we must find complex numbers ω' and ω''' (whose ratio is imaginary) satisfying the equations

$$g'' = 60 \sum{}' \frac{1}{(2m\omega' + 2n\omega''')^4}, \qquad g''' = 140 \sum{}' \frac{1}{(2m\omega' + 2n\omega''')^6}.$$

In fact, the function $\wp(z) = \wp(z \mid \omega', \omega''')$ will then obviously have the invariants g'' and g''', and hence $\wp(z)$ will satisfy the differential equation (5.30).

We now solve the inversion problem for the simplest case, where g'' and g''' are real:[13]

THEOREM 5.9. *The elliptic integral* (5.31) *can be inverted if* g'' *and* g''' *are real, with positive discriminant*[14]

$$D = \tfrac{1}{16}(g''^3 - 27g'''^2) > 0. \tag{5.32}$$

Proof. The condition (5.32) implies that the roots of the equation

$$4t^3 - g''t - g''' = 0$$

are real and distinct (recall footnote 10, p. 170). We denote these roots by e', e'' and e''', choosing the notation so that

$$e' > e'' > e'''.$$

Observing that the polynomial

$$4t^3 - g''t - g''' = 4(t - e')(t - e'')(t - e''')$$

is real for real t, and moreover positive if $t > e'$ and negative if $t < e'''$, we set[15]

$$\alpha = \int_{e'}^{+\infty} \frac{dt}{\sqrt{4t^3 - g''t - g'''}}, \tag{5.33}$$

$$\beta = \int_{-\infty}^{e'''} \frac{dt}{\sqrt{-4t^3 + g''t + g'''}}. \tag{5.34}$$

Since α and β are both positive, we can form the function $\wp(z) = \wp(z \mid \alpha, i\beta)$ figuring in Theorems 5.5 and 5.6. Let g_2 and g_3 be the invariants of $\wp(z)$. Then the rest of the proof consists in showing that $g_2 = g''$, $g_3 = g'''$. This will be done in two steps.

---

[13] However, this is the most important case in practice.

[14] The case of negative D is treated in Theorem 5.9′ below.

[15] As usual, $\sqrt{a}$ means the *positive* square root of a if $a > 0$.

Step 1. The coefficients g'' and g''' figuring in (5.33) *and* (5.34) *are uniquely determined by the numbers α and β.* First we write (5.33) and (5.34) in the form

$$\alpha = \int_{e'}^{+\infty} \frac{dt}{2\sqrt{(t-e')(t-e'')(t-e''')}}, \tag{5.35}$$

$$\beta = \int_{-\infty}^{e'''} \frac{dt}{2\sqrt{-(t-e')(t-e'')(t-e''')}}. \tag{5.36}$$

Then, making the substitution

$$t = e''' + \frac{e' - e'''}{\tau^2} \qquad (\tau > 0)$$

in (5.35) and the substitution

$$t = e' - \frac{e' - e'''}{\tau^2} \qquad (\tau > 0)$$

in (5.36), we find after a little algebraic manipulation that

$$\alpha = \frac{1}{\sqrt{e' - e'''}} \int_0^1 \frac{d\tau}{\sqrt{(1-\tau^2)(1-k^2\tau^2)}},$$

$$\beta = \frac{1}{\sqrt{e' - e'''}} \int_0^1 \frac{d\tau}{\sqrt{(1-\tau^2)(1-k'^2\tau^2)}},$$

where

$$0 < k^2 = \frac{e'' - e'''}{e' - e'''} < 1,$$

$$0 < k'^2 = \frac{e' - e''}{e' - e'''} = 1 - k^2 < 1.$$

It follows that

$$\frac{\alpha}{\beta} = \frac{\displaystyle\int_0^1 \frac{d\tau}{\sqrt{(1-\tau^2)(1-k^2\tau^2)}}}{\displaystyle\int_0^1 \frac{d\tau}{\sqrt{(1-\tau^2)(1-k'^2\tau^2)}}}.$$

As k increases from 0 to 1, the integral in the numerator increases from

$$\int_0^1 \frac{d\tau}{\sqrt{1-\tau^2}} = \frac{\pi}{2}$$

to ∞. At the same time, $k'^2 = 1 - k^2$ decreases from 1 to 0, so that the denominator decreases from ∞ to $\pi/2$. Therefore, as k^2 increases from 0 to 1, the ratio α/β increases from 0 to ∞, passing through all positive values. In other words, there is a unique value of k^2 in the interval (0, 1) corresponding to any preassigned value of the ratio α/β. Starting from

α and β, we first determine this value of k^2, and then find $e' - e'''$ from the formula

$$\sqrt{e' - e'''} = \frac{1}{\alpha}\int_0^1 \frac{d\tau}{\sqrt{(1-\tau^2)(1-k^2\tau^2)}}.$$

Next we use

$$\begin{aligned} e'' - e''' &= (e' - e''')k^2, \\ e' + e'' + e''' &= 0 \end{aligned}$$

to determine e', e'' and e''' separately, and finally

$$\begin{aligned} g'' &= -4(e'e'' + e''e''' + e'''e'), \\ g''' &= 4e'e''e''' \end{aligned} \tag{5.37}$$

to determine g'' and g''' [recall (5.24)].

Step 2. The numbers α and β satisfy the same formulas (5.33) *and* (5.34) *with g'', g''', e', e''' replaced by g_2, g_3, e_1, e_3, respectively.* According to Figure 5.4(a), the function $\wp(x) = \wp(x \mid \alpha, i\beta)$ is real and decreases from $+\infty$ to e_1 as x increases from 0 to α, while at the same time, the derivative $\wp'(x)$ increases from $-\infty$ to 0. Therefore

$$\wp'(x) = -\sqrt{4\wp^3(x) - g_2\wp(x) - g_3} \qquad (0 < x \leqslant \alpha),$$

and hence

$$\alpha = \int_{e_1}^{+\infty} \frac{dt}{\sqrt{4t^3 - g_2 t - g_3}},$$

which is the same as (5.33) with g'', g''', e' replaced by g_2, g_3, e_1. Similarly, according to Figure 5.4(d), the function $\wp(iy) = \wp(iy \mid \alpha, i\beta)$ is real and increases from $-\infty$ to e_3 as y increases from 0 to β, while at the same time, the derivative $(d/dy)\wp(iy)$ decreases from $+\infty$ to 0. Therefore

$$\frac{d}{dy}\wp(iy) = i\wp'(iy) = i(-i)\sqrt{-4\wp^3(iy) + g_2\wp(iy) + g_3} \qquad (0 < y \leqslant \beta),$$

and hence

$$\beta = \int_{-\infty}^{e_3} \frac{dt}{\sqrt{-4t^3 + g_2 t + g_3}},$$

which is the same as (5.34) with g'', g''', e''' replaced by g_2, g_3, e_3.

The proof is now an immediate consequence of Steps 1 and 2.

THEOREM 5.9′. *The elliptic integral* (5.31) *can be inverted if g'' and g''' are real, with negative discriminant*

$$D = \tfrac{1}{16}(g''^3 - 27g'''^2) < 0. \tag{5.38}$$

Proof. The condition (5.38) implies that one of the roots of the equation

$$4t^3 - g''t - g''' = 0$$

is real, while the other two are complex conjugates (again recall footnote 10, p. 170). We denote these roots by e', e'' and e''', choosing the notation so that e'' is real. Observing that the polynomial

$$4t^3 - g''t - g''' = 4(t - e')(t - e'')(t - e''')$$

is real for real t, and moreover positive if $t > e''$ and negative if $t < e''$, we set

$$2\alpha = \int_{e''}^{+\infty} \frac{dt}{\sqrt{4t^3 - g''t - g'''}}, \tag{5.33'}$$

$$2\beta = \int_{-\infty}^{e''} \frac{dt}{\sqrt{-4t^3 + g''t - g'''}}. \tag{5.34'}$$

Since α and β are both positive, we can form the function $\wp(z) = \wp(z \mid \alpha - i\beta, \alpha + i\beta)$ figuring in Theorem 5.6′. Let g_2 and g_3 be the invariants of $\wp(z)$. Then the rest of the proof consists in showing that $g_2 = g''$, $g_3 = g'''$. This will be done in two steps.

Step 1. The coefficients g'' and g''' figuring in (5.33′) *and* (5.34′) *are uniquely determined by the numbers α and β.* First we write (5.33′) and (5.34′) in the form

$$2\alpha = \int_{e''}^{+\infty} \frac{dt}{2\sqrt{(t - e')(t - e'')(t - e''')}}, \tag{5.35'}$$

$$2\beta = \int_{-\infty}^{e''} \frac{dt}{2\sqrt{-(t - e')(t - e'')(t - e''')}}. \tag{5.36'}$$

Then, making the substitution

$$t = e'' + \tau^2 \qquad (\tau > 0)$$

in (5.35′) and the substitution

$$t = e'' - \tau^2 \qquad (\tau > 0)$$

in (5.36′), we find after a little algebraic manipulation that

$$2\alpha = \int_0^{\infty} \frac{d\tau}{\sqrt{\tau^4 + 2\rho\tau^2 \cos\varphi + \rho^2}} = \sqrt{\rho} \int_0^{\infty} \frac{dt}{\sqrt{t^4 + 2t^2 \cos\varphi + 1}},$$

$$2\beta = \int_0^{\infty} \frac{d\tau}{\sqrt{\tau^4 - 2\rho\tau^2 \cos\varphi + \rho^2}} = \sqrt{\rho} \int_0^{\infty} \frac{dt}{\sqrt{t^4 - 2t^2 \cos\varphi + 1}},$$

where

$$e'' - e' = \rho e^{i\varphi}, \quad e'' - e''' = \rho e^{-i\varphi} \qquad (\rho > 0, 0 < \varphi < \pi). \tag{5.39}$$

It follows that

$$\frac{\alpha}{\beta} = \frac{\displaystyle\int_0^\infty \frac{dt}{\sqrt{t^4 + 2t^2\cos\varphi + 1}}}{\displaystyle\int_0^\infty \frac{dt}{\sqrt{t^4 - 2t^2\cos\varphi + 1}}}.$$

As φ increases from 0 to π, $\cos\varphi$ decreases from 1 to -1, and hence the integral in the numerator increases from

$$\int_0^\infty \frac{dt}{t^2 + 1} = \frac{\pi}{2}$$

to ∞. At the same time, the denominator decreases from ∞ to $\pi/2$. Therefore, as φ increases from 0 to π, the ratio α/β increases from 0 to ∞, passing through all positive values. In other words, there is a unique value of φ in the interval $(0, \pi)$ corresponding to any preassigned value of the ratio α/β. Starting from α and β, we first determine this value of φ, and then find ρ from the formula

$$\sqrt{\rho} = \frac{1}{2\alpha}\int_0^\infty \frac{dt}{\sqrt{t^4 + 2t^2\cos\varphi + 1}}.$$

Then we use (5.39) and

$$e' + e'' + e''' = 0$$

to determine e', e'' and e''' separately, and finally (5.37) to determine g'' and g'''.

Step 2. The numbers α and β satisfy the same formulas (5.33′) *and* (5.34′) *with g'', g''', e'' replaced by g_2, g_3, e_2, respectively.* According to Figure 5.6(a), the function $\wp(x) = \wp(x \mid \alpha - i\beta, \alpha + i\beta)$ is real and decreases from $+\infty$ to e_2 as x increases from 0 to 2α, while at the same time, the derivative $\wp'(x)$ increases from $-\infty$ to 0. Therefore

$$\wp'(x) = -\sqrt{4\wp^3(x) - g_2\wp(x) - g_3} \qquad (0 < x \leqslant 2\alpha),$$

and hence

$$2\alpha = \int_{e_2}^{+\infty} \frac{dt}{\sqrt{4t^3 - g_2 t - g_3}},$$

which is the same as (5.33′) with g'', g''', e'' replaced by g_2, g_3, e_2. Similarly, according to Figure 5.6(b), the function

$$\wp(2\alpha + iy) = \wp(2\alpha + iy \mid \alpha - i\beta, \alpha + i\beta)$$

is real and decreases from e_2 to $-\infty$ as y increases from 0 to 2β, while at the same time, the derivative $(d/dy)\wp(2\alpha + iy)$ decreases from 0 to $-\infty$. Therefore

$$\frac{d}{dy}\wp(2\alpha + iy) = i\wp'(2\alpha + iy) = i^2\sqrt{-4\wp^3(2\alpha + iy) + g_2\wp(2\alpha + iy) + g_3} \qquad (0 \leqslant y < \beta),$$

and hence

$$2\beta = \int_{-\infty}^{e_2} \frac{dt}{\sqrt{-4t^3 + g_2 t + g_3}},$$

which is the same as (5.34′) with g'', g''', e'' replaced by g_2, g_3, e_2.

The proof is now an immediate consequence of Steps 1 and 2.

25. The Functions $\zeta(z)$ and $\sigma(z)$

The doubly periodic function $\wp(z)$, which has a double pole with principal part

$$\frac{1}{(z - \Omega)^2}$$

at every period $\Omega = 2m\omega_1 + 2n\omega_3$, reminds us of the simply periodic function $\csc^2 z$, which has a double pole with principal part

$$\frac{1}{(z - \omega)^2}$$

at every period $\omega = m\pi$ (see Problem 5.7). But $\csc^2 z$ is closely related to other, simpler trigonometric functions, i.e., to the function $\cot z$ which has a simple pole with principal part

$$\frac{1}{z - \omega}$$

at every period ω and to the function $\sin z$ which is entire with a simple zero at every ω. In fact, apart from algebraic relations of no interest in the present context, we have

$$\frac{d}{dz} \operatorname{Ln} \sin z = \cot z, \qquad \frac{d}{dz} \cot z = -\csc^2 z.$$

This suggests the possibility of finding functions which bear the same relation to $\wp(z)$ as $\cot z$ and $\sin z$ bear to $\csc^2 z$. Although they are not elliptic themselves,[16] such functions can actually be found and are of great importance in calculations involving elliptic functions.

We begin by studying *Weierstrass' zeta function* $\zeta(z)$,[17] the analogue of $\cot z$. This function is defined by the conditions

$$\frac{d}{dz}\zeta(z) = -\wp(z), \qquad \lim_{z \to 0} \left[\zeta(z) - \frac{1}{z}\right] = 0, \tag{5.40}$$

---

[16] According to Theorem 4.17, a function whose only poles are simple poles at the periods Ω, or an entire function (with simple zeros at Ω) cannot be elliptic.

[17] Not to be confused with Riemann's zeta function appearing in Problems I.18.1, II.1.15, II.1.16 and II.10.18.

which together imply

$$\zeta(z) - \frac{1}{z} = -\int_0^z \left[\wp(Z) - \frac{1}{Z^2}\right] dZ,$$

where the integral is along any rectifiable curve passing through no periods $\Omega \neq 0$. Replacing $\wp(z)$ by its uniformly convergent partial fraction expansion (5.9) and integrating term by term, we obtain

$$\zeta(z) - \frac{1}{z} = \int_0^z \sum{}' \left[\frac{1}{(Z-\Omega)^2} - \frac{1}{\Omega^2}\right] dZ = \sum{}' \left[\frac{1}{z-\Omega} + \frac{1}{\Omega} + \frac{z}{\Omega^2}\right],$$

i.e.,

$$\zeta(z) = \frac{1}{z} + \sum{}' \left[\frac{1}{z-\Omega} + \frac{1}{\Omega} + \frac{1}{\Omega^2}\right]. \tag{5.41}$$

Thus $\zeta(z)$ is meromorphic with simple poles and corresponding principal parts

$$\frac{1}{z-\Omega}$$

at the points Ω. It is easy to see that $\zeta(z)$ is an odd function. In fact,

$$\frac{d}{dz}[\zeta(z) + \zeta(-z)] = \zeta'(z) - \zeta'(-z) = -\wp(z) + \wp(-z) = 0$$

for all z, and hence

$$\zeta(z) + \zeta(-z) = C$$

or

$$\left[\zeta(z) - \frac{1}{z}\right] + \left[\zeta(-z) + \frac{1}{z}\right] = C,$$

where C is a constant. Taking the limit of the left-hand side as $z \to 0$, we find that $C = 0$ [recall (5.40)], so that

$$\zeta(-z) = -\zeta(z),$$

as asserted.

THEOREM 5.10. *The function $\zeta(z)$ is quasi-periodic in the sense that*

$$\zeta(z + 2\omega_j) = \zeta(z) + 2\eta_j \qquad (j = 1, 2, 3), \tag{5.42}$$

i.e., $\zeta(z)$ changes by the additive constant $2\eta_j$ when z changes by $2\omega_j$. The quantities ω_j and η_j $(j = 1, 2, 3)$ satisfy the formulas

$$2\omega_3\eta_1 - 2\omega_1\eta_3 = \pi i, \quad 2\omega_1\eta_2 - 2\omega_2\eta_1 = -\pi i, \quad 2\omega_2\eta_3 - 2\omega_3\eta_2 = -\pi i, \tag{5.43}$$

known as Legendre's relations.

Proof. Since

$$\frac{d}{dz}[\zeta(z + 2\omega_j) - \zeta(z)] = \zeta'(z + 2\omega_j) - \zeta'(z) = \wp(z) - \wp(z + 2\omega_j) \equiv 0,$$

we immediately deduce (5.42), where $2\eta_j$ is a constant of integration. Moreover, the result of integrating $\zeta(z)$ along the boundary γ of the parallelogram with vertices $-\omega_2$, $\omega_1 - \omega_3$, ω_2, $-\omega_1 + \omega_3$ is

$$\int_\gamma \zeta(z)\, dz = 2\pi i,$$

since $I(\gamma)$ contains a single pole $z = 0$ with residue 1. Writing out the integral in detail, we obtain

$$\left[\int_0^1 \zeta(-\omega_1 - \omega_3 + 2\omega_1 t)2\omega_1\, dt - \int_0^1 \zeta(-\omega_1 + \omega_3 + 2\omega_1 t)2\omega_1\, dt\right] + \left[\int_0^1 \zeta(\omega_1 - \omega_3 + 2\omega_3 t)2\omega_3\, dt - \int_0^1 \zeta(-\omega_1 - \omega_3 + 2\omega_3 t)2\omega_3\, dt\right] = 2\pi i,$$

which together with (5.42) implies

$$-4\omega_1\eta_3 + 4\omega_3\eta_1 = 2\pi i,$$

and gives the first of the formulas (5.43). The other two formulas follow at once from $2\omega_2 = 2\omega_1 + 2\omega_3$ and

$$\begin{aligned} 2\eta_2 &= \zeta(z + \omega_2) - \zeta(z) \\ &= [\zeta(z + 2\omega_1 + 2\omega_3) - \zeta(z + 2\omega_3)] + [\zeta(z + 2\omega_3) - \zeta(z)] \\ &= 2\eta_1 + 2\eta_3. \end{aligned}$$

Next we construct *Weierstrass' sigma function* $\sigma(z)$, the analogue of $\sin z$ in the sense described above. This function is defined by the conditions

$$\frac{d}{dz}\operatorname{Ln}\sigma(z) = \frac{\sigma'(z)}{\sigma(z)} = \zeta(z), \qquad \lim_{z\to 0}\frac{\sigma(z)}{z} = 1, \tag{5.44}$$

which together imply

$$\operatorname{Ln}\frac{\sigma(z)}{z} = \int_0^z \left[\zeta(Z) - \frac{1}{Z}\right] dZ.$$

Recalling (5.41), we find that

$$\operatorname{Ln}\frac{\sigma(z)}{z} = \sum{}' \left[\operatorname{Ln}\left(1 - \frac{z}{\Omega}\right) + \frac{z}{\Omega} + \frac{z^2}{2\Omega^2}\right],$$

and hence

$$\sigma(z) = z\exp\left\{\int_0^z \left[\zeta(Z) - \frac{1}{Z}\right] dZ\right\} = z\prod{}' \left(1 - \frac{z}{\Omega}\right)\exp\left(\frac{z}{\Omega} + \frac{z^2}{2\Omega^2}\right), \tag{5.45}$$

where the infinite product is over all $\Omega \neq 0$ (as indicated by the prime). It

follows from (5.45) that $\sigma(z)$ is an entire function with simple zeros at the points $\Omega = 2m\omega_1 + 2n\omega_3$. The function $\sigma(z)$ is odd, since

$$\sigma(-z) = -z \exp\left\{\int_0^{-z}\left[\zeta(Z) - \frac{1}{Z}\right] dZ\right\} = -z \exp\left\{-\int_0^{z}\left[\zeta(-Z) + \frac{1}{Z}\right] dZ\right\}$$
$$= -z \exp\left\{\int_0^{z}\left[\zeta(Z) - \frac{1}{Z}\right] dZ\right\} = -\sigma(z).$$

THEOREM 5.11. *When z is replaced by $z + 2\omega_j$, $\sigma(z)$ is multiplied by the factor* $\exp[i\pi + 2\eta_j(z + \omega_j)]$, *i.e.*,

$$\sigma(z + 2\omega_j) = -\sigma(z)e^{2\eta_j(z+\omega_j)} \qquad (j = 1, 2, 3) \tag{5.46}$$

for all z.

Proof. Combining (5.42) and (5.44), we find that

$$\frac{\sigma'(z + 2\omega_j)}{\sigma(z + 2\omega_j)} - \frac{\sigma'(z)}{\sigma(z)} = 2\eta_j \qquad (j = 1, 2, 3) \tag{5.47}$$

for all z. Integration of (5.47) gives

$$\operatorname{Ln}\frac{\sigma(z + 2\omega_j)}{\sigma(z)} = 2\eta_j z + c_j$$

or

$$\sigma(z + 2\omega_j) = \sigma(z)e^{2\eta_j z + c_j}. \tag{5.48}$$

Setting $z = -\omega_j$ in (5.48) and using the oddness of $\sigma(z)$, we obtain

$$-1 = e^{-2\eta_j\omega_j + c_j},$$

which together with (5.48) implies (5.46).

The next two theorems show that the functions $\zeta(z)$ and $\sigma(z)$ can be used to represent arbitrary elliptic functions in closed form:

THEOREM 5.12. *Given an elliptic function $f(z)$ of order n, with fundamental periods $2\omega_1$ and $2\omega_3$, let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be the zeros and poles of $f(z)$ in the fundamental period parallelogram, each counted a number of times equal to its order. Then*

$$f(z) = C\frac{\sigma(z - a_1)\cdots\sigma(z - a_n)}{\sigma(z - b_1)\cdots\sigma(z - b_n')}, \tag{5.49}$$

where $\sigma(z)$ is constructed from the periods $\Omega = 2m\omega_1 + 2n\omega_3$, C is a constant, and

$$b_n' = (a_1 + \cdots + a_n) - (b_1 + \cdots + b_{n-1}). \tag{5.50}$$

Proof. The function $\sigma(z - c)$ has simple zeros at the points $c + 2m\omega_1 + 2n\omega_3$ ($m, n = 0, \pm 1, \pm 2, \ldots$), and only at these points. Moreover, according to Theorem 4.19, the number

$$b_n' - b_n = (a_1 + \cdots + a_n) - (b_1 + \cdots + b_n)$$

is a period of $f(z)$, i.e., b_n' is congruent to b_n. Therefore the zeros and poles of the meromorphic function

$$\varphi(z) = \frac{\sigma(z - a_1)\cdots\sigma(z - a_n)}{\sigma(z - b_1)\cdots\sigma(z - b_n')}$$

coincide with those of $f(z)$ in the whole z-plane. But $\varphi(z)$ is doubly periodic, with fundamental periods $2\omega_1$ and $2\omega_3$. In fact, according to (5.46),

$$\begin{aligned}\varphi(z + 2\omega_j) &= \exp\{2\eta_j[(z - a_1 - \omega_j) + \cdots + (z - a_n - \omega_j) \\ &\qquad - (z - b_1 - \omega_j) - \cdots - (z - b_n' - \omega_j)]\}\varphi(z) \\ &= \exp\{2\eta_j[(b_1 + \cdots + b_n') \\ &\qquad - (a_1 + \cdots + a_n)]\}\varphi(z) = \varphi(z),\end{aligned}$$

where (5.50) is used in the last step. Therefore $f(z)/\varphi(z)$ is an elliptic function with no poles (or zeros), and hence, by Theorem 4.15,

$$\frac{f(z)}{\varphi(z)} \equiv C = \text{const},$$

which is equivalent to (5.49).

COROLLARY. *The points $a_1, \ldots, a_n$ and $b_1, \ldots, b_n'$ in Theorem* 5.12 *can be replaced by any other congruent points satisfying the condition* (5.50).

THEOREM 5.13. *Given an elliptic function $f(z)$ with fundamental periods $2\omega_1$ and $2\omega_3$, let $b_1, \ldots, b_r$ be the poles of $f(z)$ in the fundamental period parallelogram. Suppose b_k is of order β_k, with principal part*

$$G_k(z) = \frac{A_1^{(k)}}{z - b_k} + \cdots + \frac{A_{\beta_k}^{(k)}}{(z - b_k)^{\beta_k}} \qquad (k = 1, \ldots, r).$$

Then

$$\begin{aligned}f(z) = C + \sum_{k=1}^{r} \Big\{ &A_1^{(k)}\zeta(z - b_k) - \frac{A_2^{(k)}}{1!}\zeta'(z - b_k) + \cdots \\ &+ (-1)^{\beta_k - 1}\frac{A_{\beta_k}^{(k)}}{(\beta_k - 1)!}\zeta^{(\beta_k - 1)}(z - b_k)\Big\},\end{aligned} \tag{5.51}$$

where $\zeta(z)$ is constructed from the periods $\Omega = 2m\omega_1 + 2n\omega_3$ and C is a constant.[18]

Proof. The function $\zeta^{(j)}(z - c)$, $j \geqslant 0$ has a pole of order $j + 1$ with principal part

$$(-1)^j \frac{j!}{(z - c - 2m\omega_1 - 2n\omega_3)^{j+1}}$$

[18] If $j \geqslant 1$, we can replace $\zeta^{(j)}(z - b_k)$ by $-\wp^{(j-1)}(z - b_k)$ in (5.51).

at every point $c + 2m\omega_1 + 2n\omega_3$ ($m, n = 0, \pm 1, \pm 2, \ldots$), but no other poles. Therefore the poles of the meromorphic function

$$\varphi(z) = \sum_{k=1}^{r} \left\{ A_1^{(k)}\zeta(z - b_k) + \frac{A_2^{(k)}}{1!}\zeta'(z - b_k) + \cdots + (-1)^{\beta_k - 1}\frac{A_{\beta_k}^{(k)}}{(\beta_k - 1)!}\zeta^{(\beta_k - 1)}(z - b_k) \right\}$$

coincide with those of $f(z)$ in the whole z-plane, and moreover have the same orders and principal parts. But $\varphi(z)$ is doubly periodic, with fundamental periods $2\omega_1$ and $2\omega_3$. In fact, according to (5.42),

$$\begin{aligned}\varphi(z + 2\omega_j) - \varphi(z) &= \sum_{k=1}^{r} A_1^{(k)}[\zeta(z + 2\omega_j - b_k) - \zeta(z - b_k)] \\ &= 2\eta_j \sum_{k=1}^{r} A_1^{(k)} = 0,\end{aligned}$$

where the last sum on the right vanishes by Theorem 4.16, being the sum of the residues of the elliptic function $f(z)$ at all its poles in the fundamental period parallelogram. Therefore $f(z) - \varphi(z)$ is an elliptic function with no poles, and hence, by Theorem 4.15,

$$f(z) - \varphi(z) \equiv C = \text{const},$$

which is equivalent to (5.51).

Remark. If $f(z)$ is a rational function of order n, with zeros $a_1, \ldots, a_n$ and poles $b_1, \ldots, b_n$ (each counted a number of times equal to its order), then $f(z)$ can be written as a ratio of two polynomials, each a product of n linear factors:

$$f(z) = C\frac{(z - a_1)\cdots(z - a_n)}{(z - b_1)\cdots(z - b_n)}.$$

Similarly, if $f(z)$ is a rational function with poles $b_1, \ldots, b_r$ of orders $\beta_1, \ldots, \beta_r$, then $f(z)$ has a partial fraction expansion of the form [19]

$$f(z) = C + \sum_{k=1}^{r} \left[\frac{A_1^{(k)}}{z - b_k} + \cdots + \frac{A_{\beta_k}^{(k)}}{(z - b_k)^{\beta_k}} \right].$$

Theorems 5.12 and 5.13 generalize these familiar results to the case of elliptic functions, with $\sigma(z)$ and $\zeta(z)$ playing the roles of z and $1/z$, respectively.

26. The Addition Theorem for $\wp(z)$

LEMMA. *If Z is not a period of $\wp(z)$, then*

$$\wp(z) - \wp(Z) = -\frac{\sigma(z - Z)\sigma(z + Z)}{\sigma^2(z)\sigma^2(Z)}. \tag{5.52}$$

[19] See e.g., G. Birkhoff and S. MacLane, *op. cit.*, Chap. 3, Sec. 11.

Proof. The function

$$f(z) = \wp(z) - \wp(Z)$$

is an elliptic function of order 2, whose zeros are Z, $-Z$ and all points congruent to Z and $-Z$.[20] Observing that $f(z)$ has a double pole at the origin, we apply Theorem 5.12, setting

$$a_1 = Z, \quad a_2 = -Z, \quad b_1 = 0, \quad b_2' = 0,$$

a choice which obviously satisfies the condition (5.50) [recall the corollary on p. 182]. This gives

$$f(z) = \wp(z) - \wp(Z) = C\frac{\sigma(z - Z)\sigma(z + Z)}{\sigma^2(z)}, \tag{5.53}$$

where C is a constant. To find C, we multiply both sides of (5.53) by z^2 and take the limit as $z \to 0$, obtaining

$$1 = C\frac{\sigma(-Z)\sigma(Z)}{\left[\lim_{z\to 0}\dfrac{\sigma(z)}{z}\right]^2} = -C\sigma^2(Z)$$

or

$$C = -\frac{1}{\sigma^2(Z)}.$$

With this value of C, (5.53) reduces to (5.52).

DEFINITION. *The functions*

$$\sigma_j(z) = \frac{\sigma(z + \omega_j)}{\sigma(\omega_j)}\, e^{-\eta_j z} \qquad (j = 1, 2, 3)$$

are known as associated sigma functions.

THEOREM 5.14. *In terms of the associated sigma functions,*

$$\wp'(z) = -2\frac{\sigma_1(z)\sigma_2(z)\sigma_3(z)}{\sigma^3(z)}. \tag{5.54}$$

Proof. Setting $Z = \omega_j$ in (5.51) and noting that then $\wp(Z) = e_j$, we obtain

$$\wp(z) - e_j = -\frac{\sigma(z - \omega_j)\sigma(z + \omega_j)}{\sigma^2(z)\sigma^2(\omega_j)}. \tag{5.55}$$

But according to (5.46),

$$\sigma(z + \omega_j) = -\sigma(z - \omega_j)e^{2\eta_j z},$$

which implies

$$\sigma(z - \omega_j) = -\sigma(z + \omega_j)e^{-2\eta_j z}.$$

[20] Note that this is true whether or not Z and $-Z$ are themselves congruent. If Z and $-Z$ are congruent, then $\wp'(Z) = 0$ and $f(z)$ has double zeros at Z, $-Z$ and all points congruent to Z (and $-Z$).

Substitution of this expression into (5.55) gives[21]

$$\wp(z) - e_j = \frac{\sigma^2(z + \omega_j)}{\sigma^2(z)\sigma^2(\omega_j)} e^{-2\eta_j z} = \frac{\sigma_j^2(z)}{\sigma^2(z)} \qquad (j = 1, 2, 3), \quad (5.56)$$

and hence

$$[\wp'(z)]^2 = 4[\wp(z) - e_1][\wp(z) - e_2][\wp(z) - e_3] = 4\frac{\sigma_1^2(z)\sigma_2^2(z)\sigma_3^2(z)}{\sigma^6(z)}$$

or

$$\wp'(z) = \pm 2\frac{\sigma_1(z)\sigma_2(z)\sigma_3(z)}{\sigma^3(z)}. \quad (5.57)$$

To choose the proper sign, we multiply both sides of (5.57) by z^3 and take the limit as $z \to 0$. Since

$$\lim_{z\to 0} z^3\wp'(z) = -2, \qquad \lim_{z\to 0} \frac{\sigma(z)}{z} = 1,$$

and moreover

$$\sigma_j(0) = 1 \qquad (j = 1, 2, 3),$$

we find that the minus sign should be chosen in (5.37), which then immediately gives (5.54).

THEOREM 5.15. *$\wp(z)$ satisfies the addition theorem*

$$\wp(z + Z) + \wp(z) + \wp(Z) = \frac{1}{4}\left[\frac{\wp'(z) - \wp'(Z)}{\wp(z) - \wp(Z)}\right]^2. \quad (5.58)$$

Proof. Taking the logarithmic derivative with respect to z of both sides of (5.52), we obtain

$$\frac{\wp'(z)}{\wp(z) - \wp(Z)} = \zeta(z + Z) + \zeta(z - Z) - 2\zeta(z). \quad (5.59)$$

Interchanging z and Z then gives

$$\frac{\wp'(Z)}{\wp(Z) - \wp(z)} = \zeta(z + Z) - \zeta(z - Z) - 2\zeta(Z). \quad (5.60)$$

Adding (5.59) and (5.60), and then dividing by 2, we find that

$$\frac{1}{2}\frac{\wp'(z) - \wp'(Z)}{\wp(z) - \wp(Z)} = \zeta(z + Z) - \zeta(z) - \zeta(Z). \quad (5.61)$$

Differentiation of both sides of (5.61) then yields

$$\frac{1}{2}\frac{\wp''(z)}{\wp(z) - \wp(Z)} - \frac{1}{2}\frac{\wp'(z)[\wp'(z) - \wp'(Z)]}{[\wp(z) - \wp(Z)]^2} = -\wp(z + Z) + \wp(z). \quad (5.62)$$

---

[21] In particular, choosing the appropriate branch of the square root, we have

$$\sqrt{\wp(z) - e_j} = \frac{\sigma_j(z)}{\sigma(z)} \qquad (j = 1, 2, 3).$$

Next we again interchange z and Z, obtaining

$$\frac{1}{2}\frac{\wp''(Z)}{\wp(Z)-\wp(z)} - \frac{1}{2}\frac{\wp'(Z)[\wp'(Z)-\wp'(z)]}{[\wp(Z)-\wp(z)]^2} = -\wp(z+Z)+\wp(Z). \tag{5.63}$$

Adding (5.62) and (5.63), we find that

$$\frac{1}{2}\frac{\wp''(z)-\wp''(Z)}{\wp(z)-\wp(Z)} - \frac{1}{2}\frac{[\wp'(z)-\wp'(Z)]^2}{[\wp(z)-\wp(Z)]^2} = -2\wp(z+Z)+\wp(z)+\wp(Z). \tag{5.64}$$

But

$$[\wp'(z)]^2 = 4\wp^3(z) - g_2\wp(z) - g_3$$

implies

$$2\wp'(z)\wp''(z) = 12\wp^2(z)\wp'(z) - g_2\wp'(z)$$

or

$$\wp''(z) = 6\wp^2(z) - \tfrac{1}{2}g_2.$$

Therefore

$$\frac{1}{2}\frac{\wp''(z)-\wp''(Z)}{\wp(z)-\wp(Z)} = \frac{1}{2}\frac{6[\wp^2(z)-\wp^2(Z)]}{\wp(z)-\wp(Z)} = 3[\wp(z)+\wp(Z)]. \tag{5.65}$$

Combining (5.64) and (5.65), we finally obtain (5.58).

Remark. Together with the formulas

$$[\wp'(z)]^2 = 4\wp^3(z) - g_2\wp(z) - g_3, \quad [\wp'(Z)]^2 = 4\wp^3(Z) - g_2\wp(Z) - g_3,$$

(5.58) implies an *algebraic* relation between $\wp(z+Z)$, $\wp(z)$ and $\wp(Z)$.

27. The Spherical Pendulum

The physical system consisting of a particle of mass m attached by a "weightless" rigid wire to a fixed point O and subject to the action of gravity is called a *spherical pendulum.* An equivalent system is that consisting of a particle of mass m constrained to move under the action of gravity on the surface of a smooth sphere (with center O). As we now show, the general motion of either system can easily be expressed in terms of the functions constructed in this chapter.

Suppose the position of the particle is specified by cylindrical coordinates r, θ and z,[22] with the origin at the point O and the z-axis directed vertically upward. Then the particle has kinetic energy

$$T = \frac{m}{2}(\dot{r}^2 + r^2\dot{\theta}^2 + \dot{z}^2),$$

[22] Throughout this section, z denotes the vertical position coordinate and not the complex variable $x + iy$.

where the overdot denotes differentiation with respect to time, and potential energy

$$V = mgz,$$

where g is the acceleration due to gravity. Thus conservation of energy implies

$$T + V = \frac{m}{2}(\dot{r}^2 + r^2\dot{\theta}^2 + \dot{z}^2) + mgz = E, \tag{5.66}$$

where E is the total energy of the system. Moreover, the angular momentum of the particle about the z-axis, equal to $mr^2\dot{\theta}$, must also be conserved, since the forces acting on the particle always lie in a plane containing the z-axis and hence have no moment about the z-axis.[23] It follows that

$$r^2\dot{\theta} = C, \tag{5.67}$$

where C is a positive constant.[24] Combining (5.66) and (5.67), and noting that $r = \sqrt{l^2 - z^2}$ (see Figure 5.7), we find that

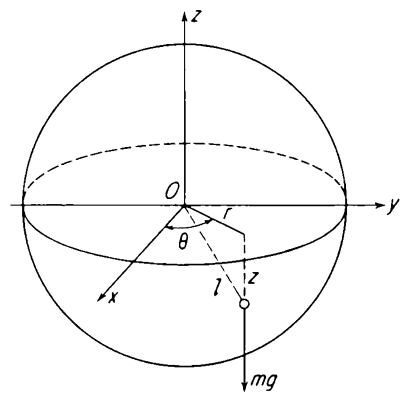


FIGURE 5.7

$$\frac{z^2\dot{z}^2}{l^2 - z^2} + \frac{C^2}{l^2 - z^2} + \dot{z}^2 = -2gz + h,$$

where

$$h = \frac{2E}{m}.$$

Therefore

$$l^2\dot{z}^2 = (h - 2gz)(l^2 - z^2) - C^2 \equiv q(z), \tag{5.68}$$

where $q(z)$ is a polynomial of degree 3.

If z_0 $(-l < z_0 < l)$ is the coordinate of the pendulum at time $t = 0$, then obviously $q(z_0) \geqslant 0$, since the initial velocity must be real. Observing that

$$q(+\infty) = +\infty, \quad q(l) = -C^2 < 0, \quad q(z_0) \geqslant 0,$$
$$q(-l) = -C^2 < 0, \quad q(-\infty) = -\infty,$$

we conclude that the zeros of $q(z)$ are all real. If $q(z_0) > 0$, each of the

---

[23] Conservation of angular momentum about the z-axis is also implied by the absence of θ (as opposed to $\dot{\theta}$) in the Lagrangian $L = T - V$. See e.g., E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge University Press, London (1964), Sec. 38.

[24] If $C = 0$, the spherical pendulum reduces to a simple pendulum, whose motion is confined to a plane.

intervals $(l, +\infty)$, (z_0, l) and $(-l, z_0)$ contains one zero. We shall confine ourselves to this case, labelling the zeros z_1, z_2, z_3 in such a way that

$$-l < z_3 < z_0 < z_2 < l < z_1. \tag{5.69}$$

Then the coordinate z satisfies the inequality

$$-l < z_3 \leqslant z \leqslant z_2 < l$$

during the entire motion of the pendulum, since $q(z)$ is nonnegative in the interval $[z_3, z_2]$ containing z_0 and changes sign in passing through the end points of $[z_3, z_2]$. As a result, the pendulum always lies between the two circles of latitude corresponding to the values z_2 and z_3.

Returning to equation (5.68), we make the change of variables

$$z = au + b,$$

where a and b are real ($a \neq 0$), obtaining

$$\dot{u}^2 = \frac{1}{a^2 l^2} q(au + b).$$

By properly choosing a and b, we can bring this equation into the form

$$\dot{u}^2 = 4u^3 - g_2 u - g_3. \tag{5.70}$$

In fact, it is only necessary to set

$$a = \frac{2l^2}{g}, \qquad b = \frac{h}{6g}.$$

Then the zeros of the polynomial $4u^3 - g_2 u - g_3$ are

$$e_j = \frac{z_j - b}{a} \qquad (j = 1, 2, 3),$$

where $e_1 > e_2 > e_3$ since $a > 0$ and $z_1 > z_2 > z_3$. Moreover, the interval $e_3 \leqslant u \leqslant e_2$ corresponds to the interval $z_3 \leqslant z \leqslant z_2$, and the discriminant $D = \frac{1}{16}(g_2^3 - 27g_3^2)$ is clearly positive. Hence, by Theorem 5.9, there exists an elliptic function $\wp(\tau) = \wp(\tau \mid \alpha, i\beta)$ satisfying the equation

$$[\wp'(\tau)]^2 = 4\wp^3(\tau) - g_2\wp(\tau) - g_3, \tag{5.71}$$

where, according to (5.33) and (5.34),

$$\alpha = \int_{e_1}^{+\infty} \frac{dx}{\sqrt{4x^3 - g_2 x - g_3}},$$

$$\beta = \int_{-\infty}^{e_3} \frac{dx}{\sqrt{-4x^3 + g_2 x + g_3}}.$$

Now let $\tau = \sigma + i\beta$, where σ is real. Then, according to Figure 5.4(c), $\wp(\sigma + i\beta)$ is a periodic function of σ with period 2α, taking all values in the

interval $[e_3, e_2]$.[25] This suggests choosing $u = \wp(\sigma + i\beta)$. Equation (5.70) then becomes

$$\left[\frac{d\wp(\sigma + i\beta)}{d\sigma}\right]^2\left(\frac{d\sigma}{dt}\right)^2 = 4\wp^3(\sigma + i\beta) - g_2\wp(\sigma + i\beta) - g_3,$$

which, because of (5.71), implies

$$\left(\frac{d\sigma}{dt}\right)^2 = 0,$$

i.e.,

$$\sigma = \pm(t - t_0),$$

where t_0 is a constant of integration. Therefore

$$u = \wp[\pm(t - t_0) + i\beta] \tag{5.72}$$

satisfies the equation of motion (5.70). Since

$$\wp(-\sigma + i\beta) = \wp(\sigma - i\beta) = \wp(\sigma + i\beta),$$

the choice of sign in (5.72) is irrelevant, and we pick the plus sign, which gives

$$u = \wp(t - t_0 + i\beta). \tag{5.73}$$

For $t = t_0$, (5.73) becomes $\wp(i\beta) = e_3 = \min u$. Measuring time from this moment, we write

$$u = \wp(t + i\beta),$$

or

$$z = au + b = a\wp(t + i\beta) + b \qquad \left(a = \frac{2l^2}{g}, \quad b = \frac{h}{6g}\right) \tag{5.74}$$

after transforming back to the variable z. Equation (5.74) gives a complete description of the behavior in time of the vertical coordinate z. In fact, z oscillates up and down between the value $z_3 = \min z$, taken at the times $0, 2\alpha, 4\alpha, \ldots$, and the value $z_2 = \max z$, taken at the times $\alpha, 3\alpha, 5\alpha, \ldots$.

We must still determine the behavior of the angular coordinate θ. It follows from (5.67) that

$$d\theta = \frac{C\,dt}{r^2} = \frac{C\,dt}{l^2 - z^2},$$

and hence

$$d\theta = \frac{C\,dt}{2l}\left(\frac{1}{z + l} - \frac{1}{z - l}\right) \tag{5.75}$$

---

[25] $\wp(\sigma + i\beta)$ increases from e_3 to e_2 as σ goes from 0 to α, and then decreases from e_2 to e_3 as σ goes from α to 2α. Because of the periodicity, $\wp(\tau)$ has the same behavior on every line $\tau = \sigma + (2k + 1)i\beta$.

According to Figures 5.4(b) and 5.4(d), p. 165, there are numbers γ and δ in the interval $(0, 2\beta)$ such that[26]

$$a\wp(\alpha + i\gamma) + b = l, \qquad a\wp(i\delta) + b = -l. \tag{5.76}$$

Moreover,

$$a^2l^2[\wp'(\alpha + i\gamma)]^2 = q(l) = -C^2,$$
$$a^2l^2[\wp'(i\delta)]^2 = q(-l) = -C^2,$$

and hence

$$\wp'(\alpha + i\gamma) = \wp'(i\delta) = \frac{iC}{al}. \tag{5.77}$$

In fact, the meromorphic function

$$a^2l^2[\wp'(\tau)]^2 - q[a\wp(\tau) + b]$$

vanishes for $\tau = t + i\beta$, where t is real, because of (5.68) and (5.74), and hence for arbitrary complex τ (why?). To justify choosing $+i$ instead of $-i$ in (5.77), we note that $\wp(\tau)$ does not change sign if $\alpha + i\gamma$ is replaced by $\alpha + i(2\beta - \gamma)$ or if $i\delta$ is replaced by $i(2\beta - \delta)$, while on the other hand, $\wp'(\tau)$ does change sign under these substitutions. Thus the interval $(0, 2\beta)$ always contains numbers γ and δ satisfying both (5.76) and (5.77).[27]

We now use (5.74), (5.76) and (5.77) to write (5.75) in the form

$$d\theta = \frac{Cdt}{2la}\left[\frac{1}{\wp(t + i\beta) - \wp(i\delta)} - \frac{1}{\wp(t + i\beta) - \wp(\alpha + i\gamma)}\right]$$
$$= \frac{dt}{2i}\left[\frac{\wp'(i\delta)}{\wp(t + i\beta) - \wp(i\delta)} - \frac{\wp'(\alpha + i\gamma)}{\wp(t + i\beta) - \wp(\alpha + i\gamma)}\right].$$

But

$$\frac{\wp'(Z)}{\wp(z) - \wp(Z)} = \zeta(z - Z) - \zeta(z + Z) + 2\zeta(Z)$$

[see (5.60)], and hence

$$d\theta = \frac{dt}{2i}[\zeta(t + i\beta - i\delta) - \zeta(t + i\beta + i\delta) + 2\zeta(i\delta) - \zeta(t - \alpha + i\beta - i\gamma)$$
$$+ \zeta(t + \alpha + i\beta + i\gamma) - 2\zeta(\alpha + i\gamma)]. \tag{5.78}$$

Integrating (5.78), we obtain

$$2i(\theta - \theta_0) = \operatorname{Ln}\sigma(t + i\beta - i\delta) - \operatorname{Ln}\sigma(t + i\beta + i\delta) + 2\zeta(i\delta)t$$
$$- \operatorname{Ln}\sigma(t - \alpha + i\beta - i\gamma) + \operatorname{Ln}\sigma(t + \alpha + i\beta + i\gamma) - 2\zeta(\alpha + i\gamma)t,$$

---

[26] Note that

$$e_2 = \frac{z_2 - b}{a} < \frac{l - b}{a} < e_1, \qquad -\infty < \frac{-l - b}{a} < \frac{z_3 - b}{a} = e_3.$$

[27] Actually it follows from Figures 5.4(b) and 5.4(d) that $\gamma \in (0, \beta)$, $\delta \in (\beta, 2\beta)$.

where θ_0 is a constant of integration. Finally, therefore,

$$e^{2i(\theta-\theta_0)} = e^{2[\zeta(i\delta)-2\zeta(\alpha+i\gamma)]t}\frac{\sigma(t+i\beta-i\delta)\sigma(t+\alpha+i\beta+i\gamma)}{\sigma(t+i\beta+i\delta)\sigma(t-\alpha+i\beta-i\gamma)}. \quad (5.79)$$

Together, (5.74) and (5.79) completely describe the motion of the spherical pendulum. (A typical trajectory is shown in Figure 5.8.) It follows from (5.46) and (5.79) that the angle θ increases by the constant amount

$$-2i\alpha[\zeta(i\delta) - \zeta(\alpha + i\gamma)] - 2i\eta_1(\alpha + i\gamma - i\delta)$$

whenever the time t increases by a period 2α.

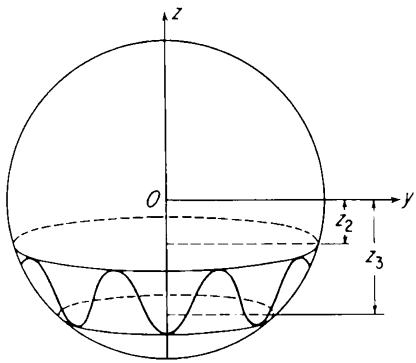


FIGURE 5.8

PROBLEMS

5.1. Is the doubly periodic function $e^{\wp(z)}$ elliptic?

5.2. Prove that every coefficient (5.15) of the Laurent expansion of $\wp(z)$ at $z = 0$ is a polynomial in the invariants g_2 and g_3, with positive rational coefficients.

5.3. Verify (5.23) directly from (5.24).

5.4. Prove that if two elliptic functions $\wp(z\,|\,\omega_1, \omega_3)$ and $\wp(z\,|\,\omega_1', \omega_3')$ have the same invariants g_2 and g_3, then ω_1, ω_3 and ω_1', ω_3' are related by a transformation of the type (4.2).

5.5. Prove that if $\wp(z)$ has real invariants, then $\wp(z)$ has fundamental periods ω_1, ω_3 which are either of the form

$$\omega_1 = \alpha, \quad \omega_3 = i\beta \qquad (\alpha, \beta > 0)$$

or of the form

$$\omega_1 = \alpha - i\beta, \quad \omega_3 = \alpha + i\beta \qquad (\alpha, \beta > 0).$$

5.6. Prove that the function $\wp(z)$ takes real values on the real axis if and only if its invariants g_2 and g_3 are real.

5.7. Verify the partial fraction expansion

$$\csc^2 z = \frac{1}{\sin^2 z} = \sum_{k=-\infty}^{\infty} \frac{1}{(z - k\pi)^2}.$$

Hint. The expansion (II.2.51) can be differentiated term by term.

5.8. Find the Laurent expansion of $\zeta(z)$ at $z = 0$. Find the power series expansion of $\sigma(z)$ at $z = 0$. In both cases, express the coefficients in terms of the invariants g_2 and g_3.

5.9. Show that Weierstrass' sigma function $\sigma(z)$ is an entire function of order 2.

Hint. Use Theorem II.10.6 and the corollary to Lemma 1, p. 156.

5.10. Generalize (5.46) to the case where $2\omega_j$ is replaced by an arbitrary period of $\wp(z)$.

5.11. Develop a theory like that of Theorems 5.12 and 5.13 for trigonometric functions, with $\sin z$, $\cos z$ playing the role of $\sigma(z)$, $\zeta(z)$, and a period strip of width 2π playing the role of the fundamental period parallelogram.

5.12. Use Theorem 5.13 to evaluate the indefinite integral

$$\int R(t, \sqrt{4t^3 - g_2 t - g_3})\, dt, \tag{5.80}$$

where $R(t, \tau)$ is a rational function, g_2 and g_3 are real, and $g_2^3 - 27g_3^2 \neq 0$.

Hint. According to Sec. 24, there exists an elliptic function $\wp(z)$ satisfying the equation

$$[\wp'(z)]^2 = 4\wp^3(z) - g_2\wp(z) - g_3.$$

Thus (5.80) becomes

$$\int R[\wp(z), \wp'(z)]\wp'(z)\, dz.$$

But $R[\wp(z), \wp'(z)]\wp'(z)$ is an elliptic function $f(z)$, with a representation of the form (5.51). Integrate (5.51) term by term, obtaining

$$\int f(z)\, dz = C_0 + Cz + \sum_{k=1}^{r} \left\{ A_1^{(k)} \operatorname{Ln} \sigma(z - b_k) - \frac{A_2^{(k)}}{1!}\zeta(z - b_k) + \cdots + (-1)^{\beta_k - 1}\frac{A_{\beta_k}^{(k)}}{(\beta_k - 1)!}\zeta^{(\beta_k - 2)}(z - b_k) \right\}.$$

5.13. Prove that any even elliptic function can be expressed as a rational function of $\wp(z)$.

5.14. Prove that any elliptic function $f(z)$ can be represented in the form

$$f(z) = R_1[\wp(z)] + R_2[\wp(z)]\wp'(z),$$

where R_1 and R_2 are rational functions.

5.15. What is the analogue of the lemma on p. 183 for rational functions and for trigonometric functions?

Ans.

$$\frac{1}{z^2} - \frac{1}{Z^2} = -\frac{(z - Z)(z + Z)}{z^2 Z^2},$$

$$\frac{1}{\sin^2 z} - \frac{1}{\sin^2 Z} = -\frac{\sin(z - Z)\sin(z + Z)}{\sin^2 z \sin^2 Z}.$$

Comment. Cf. Problem 5.11 and the remark on p. 183.

5.16. Show that

$$\wp'(z) = -\frac{\sigma(2z)}{\sigma^4(z)}.$$

Hint. Let $Z \to z$ in formula (5.52).

5.17. Find the polynomial $P(u, v, w)$ of Problem 4.6 if

$$u = \wp(z_1), \quad v = \wp(z_2), \quad w = \wp(z_1 + z_2).$$

Hint. Recall the remark on p. 186.

5.18. Prove the following *duplication formula* for $\wp(z)$:

$$\wp(2z) = \frac{1}{4}\left[\frac{\wp''(z)}{\wp'(z)}\right]^2 - 2\wp(z) \qquad (2z \text{ not a period}).$$

Hint. Let $Z \to z$ in formula (5.58).

5.19. Prove the formula

$$\wp(z + \omega_1) = e_1 + \frac{(e_1 - e_2)(e_1 - e_3)}{\wp(z) - e_1}.$$

Write similar formulas for $\wp(z + \omega_2)$ and $\wp(z + \omega_3)$.

Hint. Use Theorems 5.8 and 5.15.

5.20. Prove that the discriminant of the equation $4t^3 - g_2 t - g_3 = 0$ can be written in the form

$$\tfrac{1}{16}\wp'(z)\wp'(z + \omega_1)\wp'(z + \omega_2)\wp'(z + \omega_3),$$

where $\wp(z)$ has invariants g_2 and g_3.

Hint. Use the result of the preceding problem.

5.21. Let $q(z)$ be the same as in (5.68), and suppose $q(z_0) = 0$. Prove that one of the intervals $(-l, z_0]$ or $[z_0, l)$ contains two zeros and in fact either $z_0 = z_3$ or $z_0 = z_2$ in (5.69) [the case $z_0 = z_2 = z_3$ is also possible].

Hint. The interval $(l, +\infty)$ contains an odd number of zeros and hence just one zero, while $(-\infty, -l)$ contains an even number of zeros and hence no zeros at all.

5.22. Show that the arithmetic mean of the two levels z_2 and z_3 in the problem of the spherical pendulum always lies below the center of the sphere.

5.23. The spherical pendulum may reduce to a *conical pendulum* with the particle describing a horizontal circle at depth d below the center of the sphere. What particle velocity v is required to maintain this motion?

Ans.

$$v = \sqrt{\frac{g(l^2 - d^2)}{d}}.$$

6

ELLIPTIC FUNCTIONS: JACOBI'S THEORY

28. Jacobi's Elliptic Functions

We begin by establishing some further properties of the associated sigma functions $\sigma_j(z)$, $j = 1, 2, 3$. It follows from the definition on p. 184 and the properties of the function $\sigma(z)$ that $\sigma_j(z)$ is an entire function with simple zeros at the points

$$2m\omega_1 + 2n\omega_3 - \omega_j \qquad (j = 1, 2, 3),$$

as summarized in the following table:

$\sigma_1(z)$	$\sigma_2(z)$	$\sigma_3(z)$
$(2m - 1)\omega_1 + 2n\omega_3$	$(2m - 1)\omega_1 + (2n - 1)\omega_3$	$2m\omega_1 + (2n - 1)\omega_3$

THEOREM 6.1. *The associated sigma functions are even and satisfy the formulas*

$$\sigma_j(z + 2\omega_k) = \sigma_j(z)e^{2\eta_k(z + \omega_k)} \qquad (j \neq k), \tag{6.1}$$

$$\sigma_j(z + 2\omega_j) = -\sigma_j(z)e^{2\eta_j(z + \omega_j)}. \tag{6.2}$$

Proof. Obviously

$$\sigma_j(-z) = \frac{\sigma(-z + \omega_j)}{\sigma(\omega_j)} e^{\eta_j z} = \frac{e^{-\eta_j z}}{\sigma(\omega_j)} [-e^{2\eta_j z}\sigma(z - \omega_j)],$$

and hence, by (5.46),

$$\sigma_j(-z) = \frac{\sigma(z + \omega_j)}{\sigma(\omega_j)} e^{-\eta_j z} = \sigma_j(z),$$

which proves the evenness. [We note in passing that $\sigma_j(0) = 1$, $j = 1, 2, 3$.] Moreover, using (5.46) again, we obtain

$$\begin{aligned}\sigma_j(z + 2\omega_k) &= \exp[-\eta_j(z + 2\omega_k)] \frac{\sigma(z + \omega_j + 2\omega_k)}{\sigma(\omega_j)} \\ &= -\exp[-\eta_j(z + 2\omega_k) + 2\eta_k(z + \omega_j + \omega_k)] \frac{\sigma(z + \omega_j)}{\sigma(\omega_j)} \\ &= -\exp[2\eta_k(z + \omega_k) + 2\eta_k\omega_j - 2\eta_j\omega_k]\sigma_j(z). \qquad (6.3)\end{aligned}$$

If $j \neq k$, it follows from (5.43) that

$$2\eta_k\omega_j - 2\eta_j\omega_k = \pm\pi i,$$

which together with (6.3) implies (6.1). If $j = k$, we get (6.2) directly.

Next we introduce a special notation for certain ratios of sigma functions:

$$\lambda(z) = \frac{\sigma(z)}{\sigma_3(z)}, \quad \lambda_1(z) = \frac{\sigma_1(z)}{\sigma_3(z)}, \quad \lambda_2(z) = \frac{\sigma_2(z)}{\sigma_3(z)}. \qquad (6.4)$$

Each of the functions (6.4) is meromorphic, with simple poles at the points $2m\omega_1 + (2n - 1)\omega_3$ (and nowhere else). There zeros, all simple, are displayed in the following table:

$\lambda(z)$	$\lambda_1(z)$	$\lambda_2(z)$
$2m\omega_1 + 2n\omega_3$	$(2m - 1)\omega_1 + 2n\omega_3$	$(2m - 1)\omega_1 + (2n - 1)\omega_3$

The function $\lambda(z)$ is even, while $\lambda_1(z)$ and $\lambda_2(z)$ are odd, and moreover

$$\lambda(0) = 0, \quad \lambda'(0) = \lim_{z\to 0} \frac{\sigma(z)}{z} \frac{1}{\sigma_3(z)} = 1, \quad \lambda_1(0) = \lambda_2(0) = 1. \qquad (6.5)$$

THEOREM 6.2. *The function $\lambda(z)$ is elliptic, with fundamental periods $4\omega_1$ and $2\omega_3$.*

Proof. It follows from (5.46), (6.1) and (6.2) that

$$\lambda(z + 2\omega_1) = \frac{-\sigma(z)\exp[2\eta_1(z + \omega_1)]}{\sigma_3(z)\exp[2\eta_1(z + \omega_1)]} = -\lambda(z),$$

$$\lambda(z + 2\omega_3) = \frac{-\sigma(z)\exp[2\eta_3(z + \omega_3)]}{-\sigma_3(z)\exp[2\eta_3(z + \omega_3)]} = \lambda(z).$$

Therefore

$$\lambda(z + 4\omega_1) = -\lambda(z + 2\omega_1) = \lambda(z),$$
$$\lambda(z + 2\omega_3) = \lambda(z),$$

i.e., $\lambda(z)$ is doubly periodic (and hence elliptic, being meromorphic), with periods $4\omega_1$ and $2\omega_3$. To prove that these periods are fundamental, let z_0 be a pole and Ω a period of $\lambda(z)$. Then $z_0 + \Omega$ is also a pole, and hence

$$z_0 + \Omega = 2m'\omega_1 + (2n' - 1)\omega_3,$$

i.e.,

$$\Omega = 2\mu\omega_1 + 2\nu\omega_3,$$

where μ and ν are integers. If μ is even, Ω is a linear combination of the periods $4\omega_1$ and $2\omega_3$, with integral coefficients. However, if $\mu = 2\mu' + 1$ is odd, then

$$\Omega - 4\mu'\omega_1 - 2\nu\omega_3 = 2\omega_1,$$

so that $2\omega_1$ is a period of $\lambda(z)$. But this is impossible, since, as just shown, $\lambda(z + 2\omega_1) = -\lambda(z)$. Therefore μ can only be even, and the periods $4\omega_1$ and $2\omega_3$ are fundamental.

COROLLARY. *The function* $\lambda_1(z)$ *is elliptic with fundamental periods* $4\omega_1$ *and* $2\omega_2$, *while* $\lambda_2(z)$ *is elliptic with fundamental periods* $2\omega_1$ *and* $4\omega_3$.

Proof. Establish the relations

$$\lambda_1(z + 2\omega_1) = -\lambda_1(z), \quad \lambda_1(z + 2\omega_2) = \lambda_1(z),$$
$$\lambda_2(z + 2\omega_1) = \lambda_2(z), \quad \lambda_2(z + 2\omega_3) = -\lambda_2(z),$$

and then parallel the proof of Theorem 6.2.

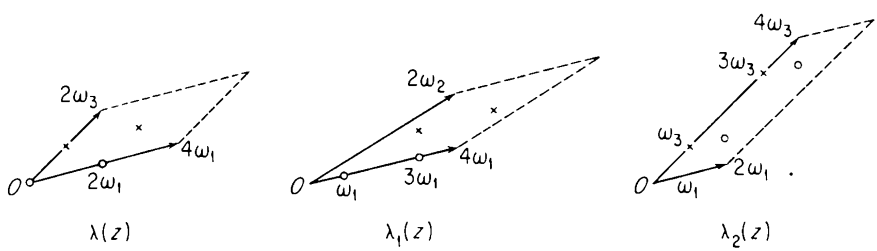


FIGURE 6.1

Figure 6.1 shows the corresponding fundamental period parallelograms of the functions (6.4). In each case, circles indicate zeros and crosses indicate poles.

Remark. Each of the functions (6.4) is an elliptic function of order 2, with two simple poles in the fundamental period parallelogram. This is to be compared with the elliptic function $\wp(z)$, also of order 2, with one double pole in the fundamental period parallelogram.

Next we introduce the notation

$$\sqrt{e_1 - e_3}\, z = u, \quad \sqrt{e_1 - e_3}\, \omega_1 = K, \quad \sqrt{e_1 - e_3}\, \omega_3 = iK', \qquad (6.6)$$

after fixing a value of the square root.[1] Here, as usual, $e_j = \wp(\omega_j)$, $j = 1, 2, 3$. We then use the functions $\lambda(z)$, $\lambda_1(z)$, $\lambda_2(z)$ to define *Jacobi's elliptic functions*

$$\begin{aligned}
\operatorname{sn} u &= \sqrt{e_1 - e_3}\,\lambda(z) = \sqrt{e_1 - e_3}\,\frac{\sigma\left(\dfrac{u}{\sqrt{e_1 - e_3}}\right)}{\sigma_3\left(\dfrac{u}{\sqrt{e_1 - e_3}}\right)}, \\
\operatorname{cn} u &= \lambda_1(z) = \frac{\sigma_1\left(\dfrac{u}{\sqrt{e_1 - e_3}}\right)}{\sigma_3\left(\dfrac{u}{\sqrt{e_1 - e_3}}\right)}, \\
\operatorname{dn} u &= \lambda_2(z) = \frac{\sigma_2\left(\dfrac{u}{\sqrt{e_1 - e_3}}\right)}{\sigma_3\left(\dfrac{u}{\sqrt{e_1 - e_3}}\right)}.
\end{aligned} \tag{6.7}$$

It follows at once from the properties of $\lambda(z)$, $\lambda_1(z)$ and $\lambda_2(z)$ that $\operatorname{sn} u$ is odd, while $\operatorname{cn} u$ and $\operatorname{dn} u$ are even, and moreover

$$\operatorname{sn}(0) = 0, \quad \operatorname{sn}'(0) = 1, \quad \operatorname{cn}(0) = 1, \quad \operatorname{dn}(0) = 1$$

[recall (6.5)].

THEOREM 6.3. *The functions* $\operatorname{sn} u$, $\operatorname{cn} u$ *and* $\operatorname{dn} u$ *are all elliptic functions, of order* 2, *with simple poles* $2mK + (2n - 1)iK'$. *The function* $\operatorname{sn} u$ *has fundamental periods* $4K$, $2iK'$ *and simple zeros* $2mK + 2niK'$, $\operatorname{cn} u$ *has fundamental periods* $4K$, $2K + 2iK'$ *and simple zeros* $(2m - 1)K + 2niK'$, *and* $\operatorname{dn} u$ *has fundamental periods* $2K$, $4iK'$ *and simple zeros* $(2m - 1)K + (2n - 1)iK'$.

Proof. Again use the established properties of the functions (6.4), together with the notation (6.7).

The behavior of the Jacobian elliptic functions for real $u = x$ is shown in Figure 6.2, where it is assumed that K and K' are also real.

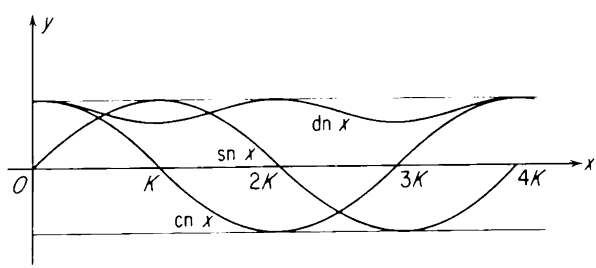


FIGURE 6.2

[1] For example, choose

$$\sqrt{e_1 - e_3} = \frac{\sigma_3(\omega_1)}{\sigma(\omega_1)}$$

(cf. footnote 21, p. 185).

Since

$$\sqrt{\wp(z) - e_j} = \frac{\sigma_j(z)}{\sigma(z)} \tag{6.8}$$

[recall (5.56) and footnote 21, p. 185], we can write sn u, cn u and dn u in terms of Weierstrass' elliptic function $\wp(z)$:

$$\operatorname{sn} u = \sqrt{\frac{e_1 - e_3}{\wp(z) - e_3}}, \quad \operatorname{cn} u = \sqrt{\frac{\wp(z) - e_1}{\wp(z) - e_3}}, \quad \operatorname{dn} u = \sqrt{\frac{\wp(z) - e_2}{\wp(z) - e_3}}. \tag{6.9}$$

It follows at once that

$$\operatorname{sn}^2 u + \operatorname{cn}^2 u = 1, \qquad k^2 \operatorname{sn}^2 u + \operatorname{dn}^2 u = 1, \tag{6.10}$$

where

$$k = \sqrt{\frac{e_2 - e_3}{e_1 - e_3}} \tag{6.11}$$

is called the *modulus* of the Jacobian elliptic functions.

THEOREM 6.4. *The derivatives of the Jacobian elliptic functions are given by the formulas*

$$(\operatorname{sn} u)' = \operatorname{cn} u \operatorname{dn} u, \tag{6.12}$$

$$(\operatorname{cn} u)' = -\operatorname{sn} u \operatorname{dn} u, \tag{6.13}$$

$$(\operatorname{dn} u)' = -k^2 \operatorname{sn} u \operatorname{dn} u. \tag{6.14}$$

Proof. The first of the formulas (6.9) implies

$$\wp(z) = e_3 + \frac{e_1 - e_3}{\operatorname{sn}^2 u},$$

and hence

$$\wp'(z) = -\frac{2(e_1 - e_3)^{3/2}(\operatorname{sn} u)'}{\operatorname{sn}^3 u}. \tag{6.15}$$

On the other hand, it follows from Theorem 5.14 that

$$\wp'(z) = -2\frac{\sigma_1(z)\sigma_2(z)}{\sigma_3(z)\sigma_3(z)}\left[\frac{\sigma(z)}{\sigma_3(z)}\right]^{-3} = -\frac{2(e_1 - e_3)^{3/2} \operatorname{cn} u \operatorname{dn} u}{\operatorname{sn}^3 u}. \tag{6.16}$$

Comparing (6.15) and (6.16), we obtain (6.12). Then, differentiating the relations (6.10) and substituting from (6.12), we immediately obtain (6.13) and (6.14).

29. Theta Functions and Their Relation to Elliptic Functions

In the theory of elliptic functions developed above, the available analytical expressions [e.g., the partial fraction expansions of $\wp(z)$ and $\zeta(z)$, and the infinite product expansion of $\sigma(z)$] do not converge rapidly enough for the

purposes of numerical computation. We now overcome this difficulty by introducing a class of periodic entire functions known as *Jacobian theta functions*. These functions will be used in Sec. 30 to write rapidly convergent expansions for the various elliptic functions.

LEMMA 1. *When z is replaced by $z + 2\omega_1$, the function*

$$f(z) = \exp\left(\frac{\eta_1 z^2}{2\omega_1} - \frac{i\pi z}{2\omega_1}\right) \tag{6.17}$$

behaves in the same way as $\sigma(z)$,[2] i.e.,

$$f(z + 2\omega_1) = -f(z)e^{2\eta_1(z+\omega_1)}.$$

Proof. The lemma follows at once by direct calculation.

LEMMA 2. *If*

$$\varphi(z) = \frac{\sigma(z)}{f(z)}, \tag{6.18}$$

where $f(z)$ is the same as in Lemma 1, then

$$\varphi(z + 2\omega_1) = \varphi(z), \qquad \varphi(z + 2\omega_3) = -\varphi(z)e^{-i\pi z/\omega_1}. \tag{6.19}$$

Proof. The first formula is an obvious consequence of Lemma 1, while the second follows from

$$\varphi(z + 2\omega_3) = -\varphi(z) \exp\left\{\frac{2\eta_3\omega_1 - 2\eta_1\omega_3}{\omega_1}(z + \omega_3) + \frac{i\pi\omega_3}{\omega_1}\right\}$$

if we use (5.43).

Remark. The function (6.18), which differs from $\sigma(z)$ by the presence of the nonzero factor $1/f(z)$, has the advantage over $\sigma(z)$ of being periodic with period $2\omega_1$.

We now expand the entire periodic function $\varphi(z)$ in a Fourier series (cf. Theorem 4.8):

THEOREM 6.5. *The function $\varphi(z)$ has the representation*

$$\varphi(z) = 2\omega_1 \frac{\vartheta_1(v)}{\vartheta_1'(0)} e^{i\pi z/2\omega_1} \qquad \left(v = \frac{z}{2\omega_1}\right), \tag{6.20}$$

in terms of the function

$$\vartheta_1(v) = i \sum_{k=-\infty}^{\infty} (-1)^k q^{(k-\frac{1}{2})^2} e^{i(2k-1)\pi v} = 2 \sum_{k=0}^{\infty} (-1)^k q^{(k+\frac{1}{2})^2} \sin(2k + 1)\pi v, \tag{6.21}$$

(called a theta function), where $q = e^{i\pi\tau}$, $\tau = \omega_3/\omega_1$. Correspondingly, $\sigma(z)$ has the representation

$$\sigma(z) = 2\omega_1 \frac{\vartheta_1(v)}{\vartheta_1'(0)} e^{\eta_1 z^2/2\omega_1}. \tag{6.22}$$

---

[2] See Theorem 5.11.

Proof. Writing

$$v = \frac{z}{2\omega_1}, \quad s = e^{i\pi v}, \quad \varphi^*(s) = \varphi\left(\frac{2\omega_1}{i\pi} \operatorname{Ln} s\right) = \varphi(z),$$

we expand the meromorphic function $\varphi^*(s)$ in powers of s^2. The resulting Laurent series

$$\varphi^*(s) = \sum_{k=-\infty}^{\infty} a_k s^{2k} \tag{6.23}$$

converges absolutely and uniformly on every compact set which does not contain the point $s = 0$ (cf. Theorem 4.7). To determine the coefficients a_k, we use the second of the formulas (6.19), which in the present notation becomes

$$\varphi^*(qs) = -s^2\varphi^*(s). \tag{6.24}$$

It follows from (6.23) and (6.24) that

$$\sum_{k=-\infty}^{\infty} a_k q^{2k} s^{2k} = -s^{-2} \sum_{k=-\infty}^{\infty} a_k s^{2k},$$

and hence

$$a_k q^{2k} = -a_{k+1} \qquad (k = 0, \pm 1, \pm 2, \ldots), \tag{6.25}$$

by the uniqueness of Laurent expansions (see Theorem II.1.2, Corollary 1).

The recurrence relation (6.25) has the solution

$$a_k = C(-1)^k q^{(k-\frac{1}{2})^2},$$

where C is a constant. Therefore

$$\varphi^*(s) = C \sum_{k=-\infty}^{\infty} (-1)^k q^{(k-\frac{1}{2})^2} s^{2k},$$

and hence

$$\begin{aligned} \varphi(z) &= C \sum_{k=-\infty}^{\infty} (-1)^k q^{(k-\frac{1}{2})^2} e^{i2k\pi v} \\ &= Ce^{i\pi v} \sum_{k=-\infty}^{\infty} (-1)^k q^{(k-\frac{1}{2})^2} e^{i(2k-1)\pi v} \\ &= -iCe^{i\pi v}\vartheta_1(v) \qquad \left(v = \frac{z}{2\omega_1}\right), \end{aligned}$$

where $\vartheta_1(v)$ is the theta function defined by (6.21). Then, according to (6.17) and (6.18),

$$\begin{aligned} \varphi(z) &= \sigma(z) \exp\left(-\frac{\eta_1 z^2}{2\omega_1} + \frac{i\pi z}{2\omega_1}\right) \\ &= -iC\vartheta_1(v)e^{i\pi z/2\omega_1}, \end{aligned} \tag{6.26}$$

i.e.,

$$\sigma(z) = -iC\vartheta_1(v)e^{\eta_1 z^2/2\omega_1}. \tag{6.27}$$

It only remains to determine the constant C. Dividing both sides of (6.27) by $z = 2\omega_1 v$ and taking the limit as $z \to 0$, we obtain

$$1 = \lim_{z\to 0} \frac{\sigma(z)}{z} = -\frac{iC}{2\omega_1} \lim_{v\to 0} \frac{\vartheta_1(v)}{v} = -\frac{iC}{2\omega_1} \vartheta_1'(0),$$

and hence

$$C = \frac{2i\omega_1}{\vartheta_1'(0)}.$$

With this value of C, (6.26) and (6.27) reduce to (6.20) and (6.22), and the theorem is proved.

Remark 1. It follows from (6.21), or indirectly from (6.22), that $\vartheta_1(v)$ is an odd periodic entire function, with period 2. Moreover $|q| < 1$, since $q = e^{i\pi\tau}$ and

$$\operatorname{Im} \tau = \operatorname{Im} \frac{\omega_3}{\omega_1} > 0.$$

Because of the presence of the factor $q^{(k-\frac{1}{2})^2}$, the convergence of the expansion (6.21) is extremely rapid. This is the decisive advantage of Jacobi's theta function over Weierstrass' sigma function.

Remark 2. Since the zeros of $\sigma(z)$ are all simple, of the form $2m\omega_1 + 2n\omega_3$, it follows from (6.22) that the zeros of $\vartheta_1(v)$ are all simple, of the form

$$v = m + n\tau \qquad (\tau = \omega_3/\omega_1;\ m, n = 0, \pm 1, \pm 2, \ldots).$$

Next we introduce three other Jacobian theta functions, related to the functions $\sigma_1(z)$, $\sigma_2(z)$ and $\sigma_3(z)$ in the same way as $\vartheta_1(v)$ is related to $\sigma(z)$. Since

$$\begin{aligned}\sigma_j(z) &= \frac{\sigma(z + \omega_j)}{\sigma(\omega_j)} e^{-\eta_j z} \\ &= \frac{2\omega_1}{\sigma(\omega_j)\vartheta_1'(0)} \vartheta_1\left(v + \frac{\omega_j}{2\omega_1}\right) \exp\left(\frac{\eta_1 z^2}{2\omega_1} + \frac{2\eta_1\omega_j - 2\omega_1\eta_j}{2\omega_1} z + \frac{\eta_1\omega_j^2}{2\omega_1}\right),\end{aligned}$$

we have

$$\sigma_1(z) = C_1\vartheta_1\left(v + \frac{1}{2}\right) e^{\eta_1 z^2/2\omega_1},$$

$$\sigma_2(z) = C_2\vartheta_1\left(v + \frac{1}{2} + \frac{\tau}{2}\right) e^{i\pi v} e^{\eta_1 z^2/2\omega_1},$$

$$\sigma_3(z) = C_3\vartheta_1\left(v + \frac{\tau}{2}\right) e^{i\pi v} e^{\eta_1 z^2/2\omega_1},$$

where C_1, C_2 and C_3 are constants. It follows from the expansion (6.21) that

$$\begin{aligned}
\vartheta_1\left(v+\frac{1}{2}\right) &= \sum_{k=-\infty}^{\infty} q^{(k-\frac{1}{2})^2} e^{i(2k-1)\pi v},\\
\vartheta_1\left(v+\frac{1}{2}+\frac{\tau}{2}\right) &= \sum_{k=-\infty}^{\infty} q^{(k-\frac{1}{2})^2} e^{i(2k-1)\pi v} e^{i(k-\frac{1}{2})\pi\tau}\\
&= q^{-1/4} e^{-i\pi v} \sum_{k=-\infty}^{\infty} q^{k^2} e^{i2k\pi v}, \qquad (6.28)\\
\vartheta_1\left(v+\frac{\tau}{2}\right) &= i \sum_{k=-\infty}^{\infty} (-1)^k q^{(k-\frac{1}{2})^2} e^{i(k-\frac{1}{2})\pi\tau}\\
&= iq^{-1/4} e^{-i\pi v} \sum_{k=-\infty}^{\infty} (-1)^k q^{k^2} e^{i2k\pi v}.
\end{aligned}$$

The remaining three theta functions are now defined as

$$\begin{aligned}
\vartheta_2(v) &= \sum_{k=-\infty}^{\infty} q^{(k-\frac{1}{2})^2} e^{i(2k-1)\pi v} = 2\sum_{k=0}^{\infty} q^{(k+\frac{1}{2})^2} \cos(2k+1)\pi v,\\
\vartheta_3(v) &= \sum_{k=-\infty}^{\infty} q^{k^2} e^{i2k\pi v} = 1 + 2\sum_{k=1}^{\infty} q^{k^2} \cos 2k\pi v, \qquad (6.29)\\
\vartheta_4(v) &= \sum_{k=-\infty}^{\infty} (-1)^k q^{k^2} e^{i2k\pi v} = 1 + 2\sum_{k=1}^{\infty} (-1)^k q^{k^2} \cos 2k\pi v,
\end{aligned}$$

and are obviously all even functions. Comparing (6.28) and (6.29), we find that

$$\begin{aligned}
\vartheta_1\left(v+\frac{1}{2}\right) &= \vartheta_2(v),\\
\vartheta_1\left(v+\frac{1}{2}+\frac{\tau}{2}\right) &= q^{-1/4} e^{-i\pi v}\, \vartheta_3(v), \qquad (6.30)\\
\vartheta_1\left(v+\frac{\tau}{2}\right) &= iq^{-1/4} e^{-i\pi v}\, \vartheta_4(v).
\end{aligned}$$

This leads to the following expressions for the associated sigma functions

$$\begin{aligned}
\sigma_1(z) &= C_1 \vartheta_2(v) e^{\eta_1 z^2/2\omega_1},\\
\sigma_2(z) &= C_2' \vartheta_3(v) e^{\eta_1 z^2/2\omega_1},\\
\sigma_3(z) &= C_3' \vartheta_4(v) e^{\eta_1 z^2/2\omega_1}.
\end{aligned}$$

To find the constants, we set $z = 0$ (and hence $v = 0$), obtaining

$$1 = C_1 \vartheta_2(0), \quad 1 = C_2' \vartheta_3(0), \quad 1 = C_3' \vartheta_4(0).$$

It follows that

$$\sigma_j(z) = \frac{\vartheta_{j+1}(v)}{\vartheta_{j+1}(0)} e^{\eta_1 z^2/2\omega_1} \qquad (j = 1, 2, 3). \qquad (6.31)$$

From the expansions (6.21) and (6.29), we can easily deduce formulas, similar to (6.30), which show how the theta functions $\vartheta_j(v)$ change when v is replaced by

$$v + \frac{1}{2}, \quad v + \frac{\tau}{2}, \quad v + \frac{1}{2} + \frac{\tau}{2}, \quad v + 1, \quad v + \tau, \quad v + 1 + \tau.$$

The results of these calculations are summarized in the following table, where $M = q^{-1/4}e^{-i\pi v}$, $N = q^{-1}e^{-2i\pi v}$:

v	$v + \frac{1}{2}$	$v + \frac{\tau}{2}$	$v + \frac{1}{2} + \frac{\tau}{2}$	$v + 1$	$v + \tau$	$v + 1 + \tau$
ϑ_1	ϑ_2	$iM\vartheta_4$	$M\vartheta_3$	$-\vartheta_1$	$-N\vartheta_1$	$N\vartheta_1$
ϑ_2	$-\vartheta_1$	$M\vartheta_3$	$-iM\vartheta_4$	$-\vartheta_2$	$N\vartheta_2$	$-N\vartheta_2$
ϑ_3	ϑ_4	$M\vartheta_2$	$iM\vartheta_1$	ϑ_3	$N\vartheta_3$	$N\vartheta_3$
ϑ_4	ϑ_3	$iM\vartheta_1$	$M\vartheta_2$	ϑ_4	$-N\vartheta_4$	$-N\vartheta_4$

It follows from (6.30) that the zeros of $\vartheta_2(v)$, $\vartheta_3(v)$ and $\vartheta_4(v)$ are all simple, and are obtained from those of $\vartheta_1(v)$ by subtracting $\frac{1}{2}$, $\frac{1}{2} + \frac{1}{2}\tau$ and $\frac{1}{2}\tau$, respectively. We summarize this information in another table:

$\vartheta_1(v)$	$\vartheta_2(v)$	$\vartheta_3(v)$	$\vartheta_4(v)$
$m + n\tau$	$m - \frac{1}{2} + n\tau$	$m - \frac{1}{2} + (n - \frac{1}{2})\tau$	$m + (n - \frac{1}{2})\tau$

In addition to being functions of v, the theta functions can also be regarded as functions of the variable τ, since the expansions (6.21) and (6.29) involve the combination $q = e^{i\pi\tau}$. This dependence on τ can be emphasized by writing $\vartheta_j(v \mid \tau)$ instead of $\vartheta_j(v)$. Clearly, each $\vartheta_j(v \mid \tau)$ is an entire function of v. Moreover, if $\operatorname{Im} \tau \geqslant \varepsilon > 0$, then

$$|q| = e^{-\pi \operatorname{Im}\tau} \leqslant e^{-\pi\varepsilon} < 1$$

for arbitrary $\varepsilon > 0$, and hence, by Theorems I.15.2 and I.15.8, each of the series (6.21) and (6.29) represents an analytic function of τ in the upper half-plane $\operatorname{Im} \tau > 0$.

THEOREM 6.6. *The theta functions satisfy the partial differential equation*

$$\frac{\partial^2 \vartheta_j(v \mid \tau)}{\partial v^2} = 4\pi i \frac{\partial \vartheta_j(v \mid \tau)}{\partial \tau} \qquad (\operatorname{Im} \tau > 0; \quad j = 1, 2, 3, 4). \quad (6.32)$$

Proof. Differentiating the uniformly convergent series (6.21) twice with respect to v and once with respect to τ, we obtain

$$\frac{\partial^2 \vartheta_1(v \mid \tau)}{\partial v^2} = -i \sum_{k=-\infty}^{\infty} (-1)^k (2k - 1)^2 \pi^2 q^{(k-\frac{1}{2})^2} e^{i(2k-1)\pi v}$$

and

$$\frac{\partial \vartheta_1(v \mid \tau)}{\partial \tau} = -\sum_{k=-\infty}^{\infty} (-1)^k (k - \tfrac{1}{2})^2 \pi q^{(k-\frac{1}{2})^2} e^{i(2k-1)\pi v},$$

which together immediately imply (6.32) for $j = 1$. The validity of (6.32) for $j = 2, 3, 4$ is verified in the same way.

We now turn to the representation of elliptic functions in terms of theta functions. It follows from (6.8), (6.22) and (6.31) that

$$\sqrt{\wp(z) - e_j} = \frac{\vartheta_1'(0)}{2\omega_1 \vartheta_{j+1}(0)} \frac{\vartheta_{j+1}(v)}{\vartheta_1(v)} \qquad (j = 1, 2, 3), \tag{6.33}$$

where $v = z/2\omega_1$. In particular, setting $z = \omega_1$, $j = 3$ and then $z = \omega_2$, $j = 3$ in (6.33), we obtain

$$\sqrt{e_1 - e_3} = \frac{\vartheta_1'(0)}{2\omega_1 \vartheta_4(0)} \frac{\vartheta_4(\frac{1}{2})}{\vartheta_1(\frac{1}{2})} = \frac{\vartheta_1'(0)}{2\omega_1 \vartheta_4(0)} \frac{\vartheta_3(0)}{\vartheta_2(0)}$$

and

$$\sqrt{e_2 - e_3} = \frac{\vartheta_1'(0)}{2\omega_1 \vartheta_4(0)} \frac{\vartheta_4(\frac{1}{2} + \frac{1}{2}\tau)}{\vartheta_1(\frac{1}{2} - \frac{1}{2}\tau)} = \frac{\vartheta_1'(0)}{2\omega_1 \vartheta_4(0)} \frac{\vartheta_2(0)}{\vartheta_3(0)}.$$

Therefore

$$k = \sqrt{\frac{e_2 - e_3}{e_1 - e_3}} = \left[\frac{\vartheta_2(0)}{\vartheta_3(0)}\right]^2 \tag{6.34}$$

[recall (6.11)]. Similarly, setting $z = \omega_1$, $j = 2$ in (6.33) gives

$$\sqrt{e_1 - e_2} = \frac{\vartheta_1'(0)}{2\omega_1 \vartheta_3(0)} \frac{\vartheta_3(\frac{1}{2})}{\vartheta_1(\frac{1}{2})} = \frac{\vartheta_1'(0)}{2\omega_1 \vartheta_3(0)} \frac{\vartheta_4(0)}{\vartheta_2(0)}.$$

Therefore, introducing the quantity

$$k' = \sqrt{\frac{e_1 - e_2}{e_1 - e_3}} \qquad (k'^2 = 1 - k^2),$$

called the *complementary modulus* of the Jacobian elliptic functions, we find that

$$k' = \left[\frac{\vartheta_4(0)}{\vartheta_3(0)}\right]^2 \tag{6.35}$$

According to (6.34) and (6.35), $\vartheta_3(0)/\vartheta_2(0)$ is one of the values of $1/\sqrt{k}$, while $\vartheta_4(0)/\vartheta_3(0)$ is one of the values of $\sqrt{k'}$. Fixing these values, we have

$$\frac{1}{\sqrt{k}} = \frac{\vartheta_3(0)}{\vartheta_2(0)}, \quad \sqrt{k'} = \frac{\vartheta_4(0)}{\vartheta_3(0)}, \quad \sqrt{\frac{k'}{k}} = \frac{\vartheta_4(0)}{\vartheta_2(0)}. \tag{6.36}$$

To obtain expressions for the Jacobian elliptic functions as ratios of theta functions, we use (6.7), (6.22) and (6.31), which immediately imply

$$\operatorname{sn} u = \sqrt{e_1 - e_3}\,\frac{2\omega_1\vartheta_4(0)}{\vartheta_1'(0)}\,\frac{\vartheta_1(v)}{\vartheta_4(v)} = \frac{\vartheta_3(0)}{\vartheta_2(0)}\,\frac{\vartheta_1(v)}{\vartheta_4(v)},$$

$$\operatorname{cn} u = \frac{\vartheta_4(0)}{\vartheta_2(0)}\,\frac{\vartheta_2(v)}{\vartheta_4(v)},$$

$$\operatorname{dn} u = \frac{\vartheta_4(0)}{\vartheta_3(0)}\,\frac{\vartheta_3(v)}{\vartheta_4(v)},$$

where

$$u = \sqrt{e_1 - e_3}\,z = 2\sqrt{e_1 - e_3}\,\omega_1 v = 2Kv$$

[recall (6.6)]. Taking account of (6.36), we can write these formulas in the particularly simple form

$$\operatorname{sn} u = \frac{1}{\sqrt{k}}\,\frac{\vartheta_1(v)}{\vartheta_4(v)}, \quad \operatorname{cn} u = \sqrt{\frac{k'}{k}}\,\frac{\vartheta_2(v)}{\vartheta_4(v)}, \quad \operatorname{dn} u = \sqrt{k'}\,\frac{\vartheta_3(v)}{\vartheta_4(v)}. \tag{6.37}$$

Remark. The prescription for using the formulas (6.37), starting from given fundamental periods $2\omega_1$ and $2\omega_3$ of the function $\wp(z)$,[3] is the following:

1. Calculate $\tau = \omega_3/\omega_1$ and then $q = e^{i\pi\tau}$.
2. Use (6.29) to find the quantities

$$\vartheta_2(0) = 2\sum_{k=0}^{\infty} q^{(k+\frac{1}{2})^2},$$

$$\vartheta_3(0) = 1 + 2\sum_{k=1}^{\infty} q^{k^2},$$

$$\vartheta_4(0) = 1 + 2\sum_{k=1}^{\infty} (-1)^k q^{2k}.$$

3. Use (6.36) to calculate $1/\sqrt{k}$, $\sqrt{k'}$ and $\sqrt{k'/k}$.
4. Use (6.21) to calculate the quantity

$$\vartheta_1'(0) = 2\sum_{k=0}^{\infty} (-1)^k(2k+1)\pi q^{(k+\frac{1}{2})^2},$$

---

[3] Note that the values of the periods $2\omega_1$, $2\omega_3$ affect sn u, cn u, dn u only through their ratio $\tau = \omega_3/\omega_1$.

and then find[4]

$$K = \sqrt{e_1 - e_3}\,\omega_1 = \frac{\vartheta_1'(0)\vartheta_3(0)}{2\vartheta_4(0)\vartheta_2(0)}.$$

Next we establish a formula which is useful in calculations involving theta functions:

THEOREM 6.7. *The theta functions satisfy the relation*

$$\vartheta_1'(0) = \pi\vartheta_2(0)\vartheta_3(0)\vartheta_4(0), \tag{6.38}$$

known as Jacobi's identity.

Proof. Using the evenness of the functions $\vartheta_2(v)$, $\vartheta_3(v)$, $\vartheta_4(v)$ and the oddness of the function $\vartheta_1(v)$, we write (6.33) in the form

$$\sqrt{\wp(2\omega_1 v) - e_j} = \frac{1}{2\omega_1}\frac{1 + \dfrac{\vartheta_{j+1}''(0)}{\vartheta_{j+1}(0)}\dfrac{v^2}{2} + \cdots}{v + \dfrac{\vartheta_1'''(0)}{\vartheta_1'(0)}\dfrac{v^3}{6} + \cdots}$$

$$= \frac{1}{2\omega_1 v}\left\{1 + \left[\frac{\vartheta_{j+1}''(0)}{\vartheta_{j+1}(0)} - \frac{1}{3}\frac{\vartheta_1'''(0)}{\vartheta_1'(0)}\right]\frac{v^2}{2} + \cdots\right\}.$$

It follows that

$$\wp(2\omega_1 v) = \frac{1}{4\omega_1 v^2} + \left[\frac{\vartheta_{j+1}''(0)}{4\omega_1^2\vartheta_{j+1}(0)} - \frac{1}{3\cdot 4\omega_1^2}\frac{\vartheta_1'''(0)}{\vartheta_1'(0)} + e_j\right] + \cdots,$$

and hence

$$e_j = \frac{1}{4\omega_1^2}\left[\frac{1}{3}\frac{\vartheta_1'''(0)}{\vartheta_1'(0)} - \frac{\vartheta_{j+1}''(0)}{\vartheta_{j+1}(0)}\right] \qquad (j = 1, 2, 3),$$

since the Laurent expansion of $\wp(2\omega_1 v)$ at the origin contains no constant term (see Theorem 5.7). Therefore

$$e_1 + e_2 + e_3 = \frac{1}{4\omega_1^2}\left[\frac{\vartheta_1'''(0)}{\vartheta_1'(0)} - \sum_{j=1}^{3}\frac{\vartheta_{j+1}''(0)}{\vartheta_{j+1}(0)}\right] = 0,$$

which implies

$$\frac{\vartheta_1'''(0)}{\vartheta_1'(0)} = \sum_{j=1}^{3}\frac{\vartheta_{j+1}''(0)}{\vartheta_{j+1}(0)}. \tag{6.39}$$

Next we set $v = 0$ in (6.32) and in the equation

$$\frac{\partial^3\vartheta_j(v\,|\,\tau)}{\partial v^3} = 4\pi i\frac{\partial^2\vartheta_j(v\,|\,\tau)}{\partial\tau\,\partial v} = 4\pi i\frac{\partial^2\vartheta_j(v\,|\,\tau)}{\partial v\,\partial\tau} \tag{6.40}$$

[4] K is needed to go from u to the new variable $v = u/2K$.

obtained from (6.32) by differentiation with respect to τ [justify the last step in (6.40)]. The result is

$$\vartheta_j''(0) = 4\pi i \frac{\partial \vartheta_j(0)}{\partial \tau}, \qquad \vartheta_j'''(0) = 4\pi i \frac{\partial \vartheta_j'(0)}{\partial \tau},$$

where we abbreviate $\vartheta(v \mid \tau)$ as $\vartheta(v)$. Therefore (6.39) takes the form

$$\frac{1}{\vartheta_1'(0)} \frac{\partial \vartheta_1'(0)}{\partial \tau} = \frac{1}{\vartheta_{j+1}(0)} \sum_{j=1}^{3} \frac{\partial \vartheta_{j+1}(0)}{\partial \tau},$$

and hence

$$\vartheta_1'(0) = C\vartheta_2(0)\vartheta_3(0)\vartheta_4(0), \tag{6.41}$$

where we must now determine the constant C. But

$$\begin{aligned} \vartheta_1'(0) &= 2\pi q^{1/4}(1 - 3q^2 + \cdots), \\ \vartheta_2(0) &= 2q^{1/4}(1 + q^2 + \cdots) \\ \vartheta_3(0) &= 1 + 2q + \cdots, \\ \vartheta_4(0) &= 1 - 2q + \cdots \end{aligned}$$

(recall the remark on p. 205), and substitution of these expansions into (6.41) gives

$$\pi(1 - 3q^2 + \cdots) = C(1 + q^2 + \cdots)(1 + 2q + \cdots)(1 - 2q + \cdots). \tag{6.42}$$

Since both sides of (6.42) must have the same constant term (why?), we find that

$$C = \pi.$$

With this value of C, (6.41) reduces to (6.38).

Equation (6.38) can be used to simplify many of the formulas appearing on p. 204. Thus

$$\sqrt{\wp(z) - e_1} = \frac{\pi\vartheta_3(0)\vartheta_4(0)}{2\omega_1} \frac{\vartheta_2(v)}{\vartheta_1(v)},$$

$$\sqrt{\wp(z) - e_2} = \frac{\pi\vartheta_2(0)\vartheta_4(0)}{2\omega_1} \frac{\vartheta_3(v)}{\vartheta_1(v)},$$

$$\sqrt{\wp(z) - e_3} = \frac{\pi\vartheta_2(0)\vartheta_3(0)}{2\omega_1} \frac{\vartheta_4(v)}{\vartheta_1(v)},$$

and

$$\sqrt{e_1 - e_3} = \frac{\pi}{2\omega_1}\vartheta_3^2(0), \quad \sqrt{e_2 - e_3} = \frac{\pi}{2\omega_1}\vartheta_2^2(0), \quad \sqrt{e_1 - e_2} = \frac{\pi}{2\omega_1}\vartheta_4^2(0).$$

Moreover, we find that

$$K = \sqrt{e_1 - e_3}\,\omega_1 = \frac{\pi}{2}\vartheta_3^2(0),$$

$$iK' = \sqrt{e_1 - e_3}\,\omega_3 = \frac{\pi\tau}{2}\vartheta_3^2(0).$$

30. Infinite Product Expansions of Theta Functions

THEOREM 6.8. *The theta functions have the infinite product expansions*

$$\begin{aligned}
\vartheta_1(v) &= 2Cq^{1/4}\sin\pi v\prod_{k=1}^{\infty}(1-2q^{2k}\cos 2\pi v+q^{4k}),\\
\vartheta_2(v) &= 2Cq^{1/4}\cos\pi v\prod_{k=1}^{\infty}(1+2q^{2k}\cos 2\pi v+q^{4k}),\\
\vartheta_3(v) &= C\prod_{k=1}^{\infty}(1+2q^{2k-1}\cos 2\pi v+q^{4k-2}),\\
\vartheta_4(v) &= C\prod_{k=1}^{\infty}(1-2q^{2k-1}\cos 2\pi v+q^{4k-2}),
\end{aligned} \tag{6.43}$$

where

$$C=\prod_{k=1}^{\infty}(1-q^{2k}). \tag{6.44}$$

Proof. Writing

$$t=e^{2i\pi v},\qquad \vartheta_3^*(t)=\vartheta_3\left(\frac{1}{2\pi i}\operatorname{Ln} t\right),$$

we expand the meromorphic function $\vartheta_3^*(t)$ in powers of t. The resulting Laurent series

$$\vartheta_3^*(t)=\sum_{k=-\infty}^{\infty}a_kt^k$$

converges absolutely and uniformly on every compact set which does not contain the point $t=0$ (cf. Theorem 4.7). Since $\vartheta_3(v)$ has simple zeros at the points $v=m-\frac{1}{2}+(n-\frac{1}{2})\tau$, the function $\vartheta_3^*(t)$ has simple zeros at the points

$$t=e^{2i\pi v}=-e^{i(2m-1)\pi\tau}=-q^{2m-1}.$$

We divide these zeros into two groups

$$-q^{-1},\quad -q^{-3},\quad -q^{-5},\ldots \tag{6.45}$$

and

$$-q,\quad -q^{3},\quad -q^{5},\ldots \tag{6.46}$$

Then the function

$$\psi(t)=\prod_{k=1}^{\infty}(1+q^{2k-1}t) \tag{6.47}$$

is entire and has the zeros (6.45), while the function

$$\chi(t)=\prod_{k=1}^{\infty}(1+q^{2k-1}t^{-1}) \tag{6.48}$$

is analytic for $0 < |t| < \infty$ and has the zeros (6.46).[5] Therefore the function

$$\psi(t)\chi(t) = \prod_{k=1}^{\infty} (1 + q^{2k-1}t)(1 + q^{2k-1}t^{-1})$$

is analytic for $0 < |t| < \infty$ and has the same zeros as $\vartheta_3^*(t)$.

Using the substitution $t = e^{2i\pi v}$ to return to the variable v, we find that

$$\mu(v) = \psi(e^{2i\pi v})\chi(e^{2i\pi v}) = \prod_{k=1}^{\infty} (1 + q^{2k-1}e^{2i\pi v})(1 + 2^{2k-1}e^{-2i\pi v})$$

is a periodic entire function of period 1, with the same zeros as the theta function $\vartheta_3(v)$. Therefore

$$\frac{\vartheta_3(v)}{\mu(v)} \tag{6.49}$$

is a periodic entire function of period 1. Moreover,

$$\begin{aligned}\mu(v + \tau) &= \prod_{k=1}^{\infty} (1 + q^{2k+1}e^{2i\pi v})(1 + q^{2k-3}e^{-2i\pi v}) \\ &= \frac{1 + q^{-1}e^{-2i\pi v}}{1 + qe^{2i\pi v}}\,\mu(v) = q^{-1}e^{-2i\pi v}\mu(v) = N\mu(v),\end{aligned}$$

where N is the same as in the table on p. 203, and hence

$$\frac{\vartheta_3(v + \tau)}{\mu(v + \tau)} = \frac{N\vartheta_3(v)}{N\mu(v)},$$

i.e., the function (6.49) also has the period τ. Therefore (6.49) is an entire elliptic function. It follows from Theorem 4.15 that

$$\frac{\vartheta_3(v)}{\mu(v)} \equiv C,$$

where C is a constant, and hence

$$\vartheta_3(v) = C \prod_{k=1}^{\infty} (1 + q^{2k-1}e^{2i\pi v})(1 + q^{2k-1}e^{-2i\pi v}).$$

---

[5] Use the convergence of the series

$$\sum_{k=1}^{\infty} |q|^{2k-1} \qquad (|q| < 1),$$

together with Theorem I.15.16 and the considerations on p. I.340, to show that each of the infinite products (6.47) and (6.48) is absolutely and uniformly convergent on every compact set [which in the case of (6.48) cannot contain the origin].

Using the third row of the indicated table, we find that

$$\vartheta_4(v) = \vartheta_3\left(v + \frac{1}{2}\right) = C\prod_{k=1}^{\infty}(1 - q^{2k-1}e^{2i\pi v})(1 - q^{2k-1}e^{-2i\pi v}),$$

$$\vartheta_2(v) = \frac{1}{M}\vartheta_3\left(v + \frac{\tau}{2}\right) = Cq^{1/4}e^{i\pi v}\prod_{k=1}^{\infty}(1 + q^{2k}e^{2i\pi v})(1 + q^{2k-2}e^{-2i\pi v})$$

$$= Cq^{1/4}(e^{i\pi v} + e^{-i\pi v})\prod_{k=1}^{\infty}(1 + q^{2k}e^{2i\pi v})(1 + q^{2k}e^{-2i\pi v}),$$

$$\vartheta_1(v) = -\frac{i}{M}\vartheta_2\left(v + \frac{1}{2} + \frac{\tau}{2}\right)$$

$$= Ciq^{1/4}e^{i\pi v}\prod_{k=1}^{\infty}(1 - q^{2k}e^{2i\pi v})(1 - q^{2k-2}e^{-2i\pi v})$$

$$= -Ciq^{1/4}(e^{i\pi v} - e^{-i\pi v})\prod_{k=1}^{\infty}(1 - q^{2k}e^{2i\pi v})(1 - q^{2k}e^{-2i\pi v}).$$

These formulas immediately imply (6.43), and it only remains to evaluate the constant C.

To find C, we use the relation (6.38). Since

$$\vartheta_1'(0) = \lim_{v\to 0}\frac{\vartheta_1(v)}{v} = 2C\pi q^{1/4}\prod_{k=1}^{\infty}(1 - q^{2k})^2,$$

$$\vartheta_2(0) = 2Cq^{1/4}\prod_{k=1}^{\infty}(1 + q^{2k})^2,$$

$$\vartheta_3(0) = C\prod_{k=1}^{\infty}(1 + q^{2k-1})^2,$$

$$\vartheta_4(0) = C\prod_{k=1}^{\infty}(1 - q^{2k-1})^2,$$

it follows from (6.38) that

$$\prod_{k=1}^{\infty}(1 - q^{2k})^2 = C^2\prod_{k=1}^{\infty}(1 - q^{2k})^2(1 + q^{2k-1})^2(1 - q^{2k-1})^2. \quad (6.50)$$

On the other hand,[6] we have

$$\prod_{k=1}^{\infty}(1 - q^k)^2 = \prod_{k=1}^{\infty}(1 - q^{2k})^2(1 - q^{2k-1})^2$$

$$= \prod_{k=1}^{\infty}(1 - q^k)^2(1 + q^k)^2(1 - q^{2k-1})^2,$$

and hence

$$1 = \prod_{k=1}^{\infty}(1 + q^k)^2(1 - q^{2k-1})^2$$

$$= \prod_{k=1}^{\infty}(1 + q^{2k})^2(1 + q^{2k-1})^2(1 - q^{2k-1})^2. \quad (6.51)$$

[6] Here, as elsewhere, we make free use of the absolute convergence (for $|q| < 1$) to rearrange the infinite products (cf. p. I.337).

Comparing (6.50) and (6.51), we find that

$$C^2 = \prod_{k=1}^{\infty} (1 - q^{2k})^2,$$

or

$$C = \pm \prod_{k=1}^{\infty} (1 - q^{2k}). \tag{6.52}$$

Regarded as a function of τ (through the combination $q = e^{i\pi\tau}$), C is analytic and nonvanishing in the upper half-plane,[7] and hence one of the two signs must be chosen in (6.52). But

$$\vartheta_3(0) = 1 = C \prod_{k=1}^{\infty} (1 + q^{2k-1})^2$$

implies that $C \to 1$ as $q \to 0$. Therefore the plus sign should be chosen in (6.52), which then reduces to (6.44), thereby completing the proof.

COROLLARY. *The Weierstrass functions* $\zeta(z)$ *and* $\wp(z)$ *have the partial fraction expansions*

$$\zeta(z) = 2\eta_1 v + \frac{\pi}{2\omega_1} \cot \pi v - \frac{i\pi}{\omega_1} \sum_{k=1}^{\infty} \left[\frac{q^{2k} e^{2i\pi v}}{1 - q^{2k} e^{2i\pi v}} - \frac{q^{2k} e^{-2i\pi v}}{1 - q^{2k} e^{-2i\pi v}} \right] \tag{6.53}$$

and

$$\wp(z) = -\frac{\eta_1}{\omega_1} + \frac{\pi^2}{4\omega_1^2} \csc^2 \pi v - \frac{\pi^2}{\omega_1^2} \sum_{k=1}^{\infty} \left[\frac{q^{2k} e^{2i\pi v}}{(1 - q^{2k} e^{2i\pi v})^2} + \frac{q^{2k} e^{-2i\pi v}}{(1 - q^{2k} e^{-2i\pi v})^2} \right], \tag{6.54}$$

where $z = 2\omega_1 v$.

Proof. Using (6.22) and the first of the formulas (6.43), we find that

$$\sigma(z) = C_1 e^{2\omega_1 \eta_1 v^2} \sin \pi v \prod_{k=1}^{\infty} (1 - q^{2k} e^{2i\pi v})(1 - q^{2k} e^{-2i\pi v}). \tag{6.55}$$

Then (6.53) and (6.54) follow immediately upon differentiating with respect to z:

$$\zeta(z) = \frac{d}{dz} \operatorname{Ln} \sigma(z), \qquad \wp(z) = -\frac{d}{dz} \zeta(z).$$

Remark. Because of the factors q^{2k}, the expansions (6.53) and (6.54) converge very rapidly. Therefore, in all numerical calculations involving Weierstrass functions, they are to be preferred to the expansions originally used to define the functions.

---

[7] Note that C is a positive real number if τ is purely imaginary.

PROBLEMS

6.1. Prove the formulas

$$\operatorname{sn}(u + 2mK + 2inK') = (-1)^m \operatorname{sn} u,$$
$$\operatorname{cn}(u + 2mK + 2inK') = (-1)^{m+n} \operatorname{cn} u,$$
$$\operatorname{dn}(u + 2mK + 2inK') = (-1)^n \operatorname{dn} u.$$

6.2. Expand sn u, cn u and dn u in power series at the point $u = 0$.

6.3. Prove the addition theorem

$$\operatorname{sn}(u + v) = \frac{\operatorname{sn} u \operatorname{cn} v \operatorname{dn} v + \operatorname{sn} v \operatorname{cn} u \operatorname{dn} u}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v}.$$

Hint. If $s_1 = \operatorname{sn} u$, $s_2 = \operatorname{sn} v$ and $u + v = \text{const}$, then

$$\frac{d}{du} \operatorname{Ln}(s_1' s_2 - s_2' s_1) = \frac{d}{du} \operatorname{Ln}(1 - k^2 s_1^2 s_2^2),$$

where the prime denotes differentiation with respect to u, and hence

$$\frac{s_1' s_2 - s_2' s_1}{1 - k^2 s_1^2 s_2^2} = \text{const}.$$

Prove that this constant equals sn $(u + v)$ and use Theorem 6.4.

6.4. Use the result of the preceding problem to prove that

$$\operatorname{cn}(u + v) = \frac{\operatorname{cn} u \operatorname{cn} v - \operatorname{sn} u \operatorname{sn} v \operatorname{dn} u \operatorname{dn} v}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v},$$

$$\operatorname{dn}(u + v) = \frac{\operatorname{dn} u \operatorname{dn} v - k^2 \operatorname{sn} u \operatorname{sn} v \operatorname{cn} u \operatorname{cn} v}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v}.$$

6.5. Prove that $z = \operatorname{sn} w$ is the inverse of the elliptic integral

$$w = f(z) = \int_0^z \frac{dt}{\sqrt{(1 - t^2)(1 - k^2 t^2)}},$$

defined on p. II.124.

6.6. Onto what domain does $z = \operatorname{sn} w$ map the rectangle with vertices $-K$, K, $K + iK'$, $-K + iK'$?

Hint. Recall Sec. II.19.

6.7. Prove that sn $K = 1$, cn $K = 0$, dn $K = k'$.

6.8. Show that

$$K = \int_0^{\pi/2} \frac{d\varphi}{\sqrt{1 - k^2 \sin^2 \varphi}}, \qquad K' = \int_0^{\pi/2} \frac{d\varphi}{\sqrt{1 - k'^2 \sin^2 \varphi}}.$$

6.9. Show that

$$K + iK' = \int_0^{1/k} \frac{dt}{\sqrt{(1 - t^2)(1 - k^2t^2)}},$$

and hence that

$$\operatorname{sn}(K + iK') = 1/k, \quad \operatorname{cn}(K + iK') = -ik'/k, \quad \operatorname{dn}(K + iK') = 0.$$

Hint. Cf. p. II.126.

6.10. Use Problems 6.3, 6.4 and 6.9 to deduce the periods of sn u, cn u and dn u. Find formulas for sn $(u + iK')$, cn $(u + iK')$ and dn $(u + iK')$.

6.11. Find the residues of sn u, cn u and dn u at $u = iK'$.

Ans. $1/k$, $-i/k$, $-i$.

6.12. Use Jacobian elliptic functions to determine the motion of the spherical pendulum of Sec. 27.

Hint. Use Problem 6.5 after making the substitutions

$$v = \sqrt{\frac{u - e_3}{e_2 - e_3}}, \quad k = \sqrt{\frac{e_2 - e_3}{e_1 - e_3}}, \quad p = \sqrt{e_1 - e_3},$$

which transform

$$\dot{u}^2 = 4u^3 - g_2u - g_3$$

into

$$\dot{v}^2 = p^2(1 - v^2)(1 - k^2v^2).$$

Ans. $z - z_3 = (z_2 - z_3)\operatorname{sn}^2 p(t - t_0)$.

6.13. Use Jacobian elliptic functions to determine the motion of a simple pendulum of mass m and length l (cf. footnote 24, p. 187).

Hint. Let θ be the angle between the pendulum and the vertical direction. Then conservation of energy gives

$$\frac{m}{2} l^2\dot{\theta}^2 - mg\, l \cos\theta = E,$$

instead of (5.66).

Ans. If α is the amplitude of the motion, i.e., the maximum value of θ, then

$$\sin\frac{\theta}{2} = \sin\frac{\alpha}{2} \operatorname{sn} p(t - t_0),$$

where

$$k = \sin\frac{\alpha}{2}, \quad p = \sqrt{\frac{g}{l}}.$$

6.14. Use Problems 6.8 and 6.13 to investigate the dependence of the period of the motion of a simple pendulum on the amplitude of the oscillations. What is the small-amplitude approximation?

6.15. Verify the table on p. 203.

6.16. Show that

$$\vartheta_4^2(0)\vartheta_2^2(z) = \vartheta_2^2(0)\vartheta_4^2(z) - \vartheta_3^2(0)\vartheta_1^2(z),$$

and write three other formulas of a similar type.

Hint. The function

$$\frac{a\vartheta_1^2(z) + b\vartheta_4^2(z)}{\vartheta_2^2(z)}$$

is an elliptic function of order 1. Use Theorem 4.17 and determine the constants a and b.

6.17. Prove that

$$\frac{\vartheta_3(z\,|\,\tau)\vartheta_4(z\,|\,\tau)}{\vartheta_4(2z\,|\,2\tau)} = \frac{\vartheta_3(0\,|\,\tau)\vartheta_4(0\,|\,\tau)}{\vartheta_4(0\,|\,2\tau)},$$

$$\frac{\vartheta_2(z\,|\,\tau)\vartheta_1(z\,|\,\tau)}{\vartheta_1(2z\,|\,2\tau)} = \frac{\vartheta_3(0\,|\,\tau)\vartheta_4(0\,|\,\tau)}{\vartheta_4(0\,|\,2\tau)}.$$

6.18. Find the relation between the constant C_1 in (6.55) and the constant C in (6.44).

Ans. $C_1 = 1/\pi C^2$.

6.19. Prove that

$$\sqrt[4]{k} = \sqrt{2}q^{1/8}\prod_{k=1}^{\infty}\frac{1+q^{2k}}{1+q^{2k-1}}.$$

Find a similar formula for $\sqrt[4]{k'}$.

Hint. Use (6.36) and (6.43).

Part 3

RIEMANN SURFACES ANALYTIC CONTINUATION

7

RIEMANN SURFACES

31. Topological Preliminaries

First we define a topological space, then a topological surface, and eventually (in Sec. 32) an abstract Riemann surface. In what follows, E denotes an infinite set, whose elements are called *points* (of E), and $\{\mathcal{N}\}$ denotes a family of subsets of E, called *neighborhoods* (of points of E). Moreover, any neighborhood containing a point $e \in E$ is called a *neighborhood of* e and is denoted by $\mathcal{N}(e)$.

DEFINITION 1. *A set E and a family of neighborhoods $\{\mathcal{N}\}$ is called a topological space if*

1. *Given any point $e \in E$, there is at least one neighborhood $\mathcal{N}(e)$ containing e;*
2. *Given any two neighborhoods $\mathcal{N}_1(e)$ and $\mathcal{N}_2(e)$ containing the same point $e \in E$, there is a third neighborhood $\mathcal{N}_3(e)$ containing e such that $\mathcal{N}_3(e) \subset \mathcal{N}_1(e)$, $\mathcal{N}_3(e) \subset \mathcal{N}_2(e)$;*
3. *Given any point $e' \in \mathcal{N}(e)$, there is a neighborhood $\mathcal{N}(e') \subset \mathcal{N}(e)$.*

Remark. By a "topological space E," we mean, more exactly, a set E together with a family of neighborhoods $\{\mathcal{N}\}$ satisfying Definition 1.

DEFINITION 2. *A topological space E is called a Hausdorff space if, given any two distinct points e, $e' \in E$, there are disjoint neighborhoods $\mathcal{N}(e)$ and $\mathcal{N}(e')$.*[1]

All the familiar topological concepts introduced in Chaps. I.3 and I.4 for point sets in the complex plane have their counterparts for an arbitrary

[1] Thus $e \in \mathcal{N}(e)$, $e' \in \mathcal{N}(e')$, $e \neq e'$, $\mathcal{N}(e) \cap \mathcal{N}(e') = 0$.

Hausdorff space E. For example, a point $e \in E$ is said to be a *limit point* of a set $M \subset E$ if every neighborhood $\mathcal{N}(e)$ contains infinitely many points of M, a set $M \subset E$ is said to be *open* if every point $e \in M$ is an *interior point* of M, i.e., has a neighborhood $\mathcal{N}(e) \subset M$, a set $M \subset E$ is said to be *connected* if given any decomposition of E into two nonempty disjoint sets E_1 and E_2, at least one of the sets E_1 and E_2 contains a limit point of the other, and so on.

DEFINITION 3. *A connected Hausdorff space E is called a two-dimensional manifold if E is locally homeomorphic to the plane (synonymously, locally Euclidean) in the following sense: Given any point $e_0 \in E$, there is a neighborhood $\mathcal{N}_{e_0}$ and a mapping $z = \varphi_{e_0}(e)$ defined on $\mathcal{N}_{e_0}$ such that*[2]

1. *$\varphi_{e_0}(e)$ is one-to-one and continuous in both directions;*[3]
2. *$\varphi_{e_0}(e)$ maps $\mathcal{N}_{e_0}$ onto a simply connected domain G_{e_0} in the z-plane.*

Remark. The mapping $z = \varphi_{e_0}(e)$ carries every open subset $M \subset \mathcal{N}_{e_0}$ into an open subset of G_{e_0}. In fact, let $z_1 = \varphi_{e_0}(e_1)$ be a point of the image $\varphi_{e_0}(M)$, and suppose z_1 is not an interior point of $\varphi_{e_0}(M)$. Then G_{e_0} contains a sequence of points $\{a_n\}$ converging to z_1 such that $a_n \notin \varphi_{e_0}(M)$ for all n. Therefore, since $\varphi_{e_0}^{-1}(z)$ is continuous, the inverse images $b_n = \varphi_{e_0}^{-1}(a_n)$ converge to e_1 and hence belong to M, starting from some value of n (since e_1 is an interior point of M). But then $a_n = \varphi_{e_0}(b_n)$ must belong to $\varphi_{e_0}(M)$ starting from some value of n, contrary to the choice of $\{a_n\}$. This contradiction shows that $\varphi_{e_0}(M)$ is itself open.

DEFINITION 4. *A two-dimensional manifold E is called a (topological) surface if it has a countable base, i.e., if there is a countable family B of open sets of E such that given any point $e_0 \in E$ and neighborhood $\mathcal{N}(e_0)$, there is a set $\mathcal{O} \in B$ satisfying the relation $e_0 \in \mathcal{O} \subset \mathcal{N}(e_0)$.*

DEFINITION 5. *A topological surface E is said to be closed (or compact) if every infinite subset of E has at least one limit point. Otherwise, E is said to be open.*[4]

We now give some examples of topological surfaces:

Example 1. The finite Euclidean plane with the usual definition of a neighborhood is an open surface, while the sphere (or extended plane) is a closed surface.

---

[2] Thus $\mathcal{N}_{e_0}$ denotes a *fixed* neighborhood of e_0 [the domain of $\varphi_{e_0}(e)$], as opposed to $\mathcal{N}(e_0)$, which denotes *any* neighborhood of e_0.

[3] Continuity of the mapping $z = \varphi_{e_0}(e)$ at a point $e_1 \in \mathcal{N}_{e_0}$ means the following: Given any neighborhood $K: |z - z_1| < \rho$ of the point $z_1 = \varphi(e_1)$, there is a neighborhood $\mathcal{N}(e_1) \subset \mathcal{N}_{e_0}$ such that $z = \varphi_{e_0}(e) \in K$ if $e \in \mathcal{N}(e_1)$. Continuity of the inverse mapping $e = \varphi_{e_0}^{-1}(z)$ is defined in the same way, and continuity in both directions means that both $z = \varphi_{e_0}(e)$ and its inverse are continuous (cf. Chap. I.6).

[4] Thus an open surface contains infinite sets with no limit points.

Example 2. Let E be the torus obtained by rotating the circle $(\xi - R)^2 + \zeta^2 = r^2$ of radius $r < R$ about the ζ-axis. Clearly, E has the parametric equations

$$\xi = (R + r\cos\phi)\cos\lambda, \quad \eta = (R + r\cos\phi)\sin\lambda, \quad \zeta = r\sin\phi, \quad (7.1)$$

where $\lambda \in [0, 2\pi)$ is the longitude and $\phi \in [0, 2\pi)$ is the latitude (see Figure 7.1). Let a neighborhood of the point $e_0 = (\xi_0, \eta_0, \zeta_0) \in E$ be defined as the

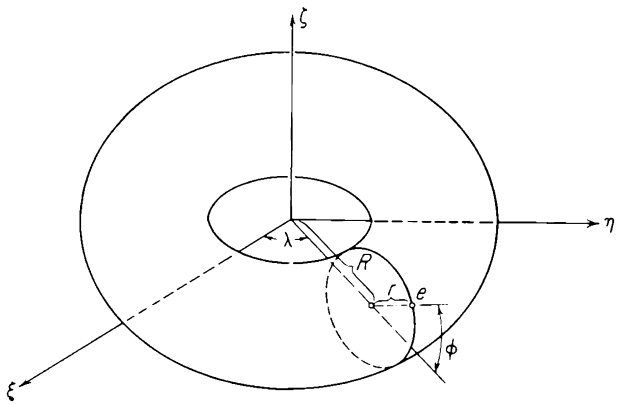


FIGURE 7.1

set $E \cap K$, where K is any open ball with center e_0 and radius not exceeding $\rho = \min\{2r, 2(R - r)\}$.[5] Then E is a closed topological surface.

Example 3. Suppose we divide the finite z-plane Π into strips

$$2k\pi \leqslant \operatorname{Re} z < 2(k + 1)\pi \qquad (k = 0, \pm 1, \pm 2, \ldots), \qquad (7.2)$$

calling two points z and z' congruent if $z' - z = 2m\pi$, where m is an integer. Then, given any point $z \in \Pi$, there are countably many points congruent to z, i.e., one in each of the strips (7.2). Letting M_z denote the class of all points congruent to z, we can regard any point $z_0 \in M_z$ as representing the whole class M_z in the sense that M_z coincides with the set of points congruent to z_0.

We now define E as the set of all classes M_z, $z \in \Pi$, and a neighborhood of M_z as the set of all classes represented by the points belonging to any open disk of radius not exceeding π with center at any point $z_0 \in M_z$. Then E is an open topological surface. In fact, E is homeomorphic to the right circular cylinder

$$\xi = \rho\cos\theta, \quad \eta = \rho\sin\theta, \qquad (0 \leqslant \theta \leqslant 2\pi, \ \zeta \text{ arbitrary}),$$

---

[5] In other words, K is any set of points (ξ, η, ζ) such that

$$(\xi - \xi_0)^2 + (\eta - \eta_0)^2 + (\zeta - \zeta_0)^2 < \rho^2,$$

i.e., the three-dimensional analogue of an open disk.

as we see by associating the point $\xi = \rho \cos x$, $\eta = \rho \sin x$, $\zeta = y$ of the cylinder with the class represented by the point $z = x + iy$ in the "fundamental strip" $0 \leqslant x < 2\pi$.

Example 4. Suppose we divide the finite z-plane Π into rectangles

$$2m\alpha \leqslant \operatorname{Re} z < 2(m+1)\alpha, \quad 2n\beta \leqslant \operatorname{Im} z < 2(n+1)\beta \qquad (m, n = 0, \pm 1, \pm 2, \ldots), \tag{7.3}$$

where α and β are fixed positive numbers, calling two points z and z' congruent if $z' - z = 2m\alpha + 2ni\beta$. Then, given any point $z \in \Pi$, there are countably many points congruent to z, i.e., one in each of the rectangles (7.3). Letting M_z denote the class of all points congruent to z, we can regard any point $z_0 \in M_z$ as representing the whole class M_z in the sense that M_z coincides with the set of points congruent to z_0.

We now define E as the set of all classes M_z, $z \in \Pi$, and a neighborhood of M_z as the set of all classes represented by the points belonging to any open disk of radius not exceeding $\min(\alpha, \beta)$ with center at any point $z_0 \in M_z$. Then E is a closed topological surface. In fact, E is homeomorphic to the torus considered in Example 2, as we see by associating the point of the torus with parameters

$$\lambda = \frac{2\pi}{\alpha} x, \qquad \phi = \frac{2\pi}{\beta} y$$

[cf. (7.1)] with the class represented by the point $z = x + iy$ in the "fundamental rectangle" $0 \leqslant x < 2\alpha$, $0 \leqslant y < 2\beta$.

Example 5. Finally consider the strip

$$-1 < \operatorname{Im} z < 1, \tag{7.4}$$

and call two points $z = x + iy$ and $z' = x' + iy'$ congruent if

$$x' - x = 4k\pi, \qquad y' = y$$

or if

$$x' - x = 2(2k+1)\pi, \qquad y' = -y$$

(where k is an integer). Then, given any point $z \in \Pi$, there are countably many points congruent to z, i.e., one in each rectangle

$$2k\pi \leqslant x < 2(k+1)\pi, \qquad -1 < y < 1.$$

Letting M_z denote the class of all points congruent to z, we can regard any point $z_0 \in M_z$ as representing the whole class M_z, just as in the preceding two examples.

Now let E be the set of all classes M_z, $z \in \Pi$, and let a neighborhood of M_z be the set of all classes represented by the points belonging to any open disk contained in the strip (7.4). Then E is an open topological surface. It is easy to see that E is homeomorphic to the *Möbius strip*, models of which

can be obtained

1. By taking a rectangular strip of paper, holding one of the (narrow) ends fixed, twisting the other through 180°, and then pasting the ends together (so that points symmetric with respect to the center of the rectangle overlap);
2. By a construction resembling that of the torus in Example 2, except that now instead of rotating the circle $C:(\xi - R)^2 + \zeta^2 = r^2$ about the ζ-axis, we rotate the radius of C making the angle $\lambda = 0$ with the $\xi\eta$-plane and introduce a new parameter $\rho \in [0, r]$ as shown in Figure 7.2.

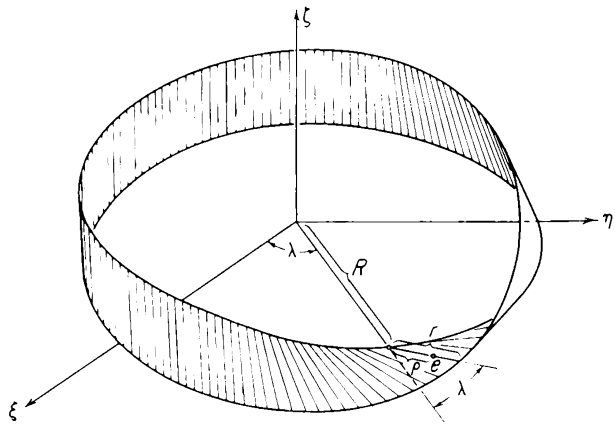


FIGURE 7.2

32. Abstract Riemann Surfaces

Let E be a topological surface, equipped with a family of neighborhoods $\{\mathcal{N}_{e_0}\}$, one for each point $e_0 \in E$, and a corresponding family of functions $\{\varphi_{e_0}(e)\}$, where $z = \varphi_{e_0}(e)$ is a homeomorphic mapping of $\mathcal{N}_{e_0}$ onto some simply connected domain G_{e_0}. Moreover, let $\varphi_{e_0}(e_0) = 0$.[6] Then $z = \varphi_{e_0}(e)$ is called the *local (uniformizing) parameter* for e_0, with corresponding *parametric neighborhood* $\mathcal{N}_{e_0}$. We are particularly interested in surfaces on which it is possible to define analytic functions, just as in the case of domains in the finite plane or in the extended plane (Riemann sphere). With this in mind, let $F(e)$ be a complex-valued function defined in the neighborhood $\mathcal{N}_{e_0}$. Then

$$f(z) = F[\varphi_{e_0}^{-1}(z)]$$

is a function of the complex variable z, defined on the domain G_{e_0}, and it is natural to call $F(e)$ analytic at a point $e_1 \in \mathcal{N}_{e_0}$ if $f(z)$ is analytic at

[6] Clearly, this entails no loss of generality.

$\zeta = \varphi_{e_0}(e_1) \in G_{e_0}$. However, e_1 also belongs to the neighborhood $\mathcal{N}_{e_1}$, and we want $F(e)$ to be analytic at e_1 regardless of which of the functions $z = \varphi_{e_0}(e)$ or $z' = \varphi_{e_1}(e)$ is used as the local parameter. But this imposes a special requirement on the surface E, as is apparent from

THEOREM 7.1. *Let E be a topological surface equipped with local parameters $\{\varphi_{e_0}(e)\}$ and parametric neighborhoods $\{\mathcal{N}_{e_0}\}$, $e_0 \in E$,[7] and let $F(e)$ be a complex-valued function defined on $\mathcal{N}_{e_0}$. Suppose*

$$f(z) = F[\varphi_{e_0}^{-1}(z)] \tag{7.5}$$

is analytic at $\zeta = \varphi_{e_0}(e_1)$. Then

$$f^*(z') = F[\varphi_{e_1}^{-1}(z')] \tag{7.6}$$

is analytic at $z' = 0$ if and only if $z = \varphi_{e_0}[\varphi_{e_1}^{-1}(z')]$ is analytic at $z' = 0$.

Proof. We assume that $e_0 \neq e_1$, since otherwise the theorem is trivial. Suppose first that (7.5) is analytic at $\zeta = \varphi_{e_0}(e_1)$ and $z = \varphi_{e_0}[\varphi_{e_1}^{-1}(z')]$ is analytic at $z' = 0$. Then[8]

$$f^*(z') = f\varphi_{e_0}\varphi_{e_1}^{-1}(z') = F\varphi_{e_0}^{-1}\varphi_{e_0}\varphi_{e_1}^{-1}(z') = F[\varphi_{e_1}^{-1}(z')]$$

is analytic at $z' = 0$, since an analytic function of an analytic function is itself analytic. Conversely, let $F(e) = \varphi_{e_0}(e)$. Then

$$f(z) = \varphi_{e_0}[\varphi_{e_0}^{-1}(z)] = z$$

is obviously analytic at ζ (in fact, in the whole z-plane), and analyticity of (7.6) at $z' = 0$ reduces at once to analyticity of $z = \varphi_{e_0}[\varphi_{e_1}^{-1}(z')]$ at $z' = 0$, thereby completing the proof.

Thus Theorem 7.1 immediately singles out a special class of topological surfaces with the structure needed to "support" analytic functions:

DEFINITION 1. *A topological surface E, equipped with local parameters $\{\varphi_{e_0}(e)\}$ and parametric neighborhoods $\{\mathcal{N}_{e_0}\}$, $e_0 \in E$, is called an abstract Riemann surface if*

$$z = \varphi_{e_0}(e), \quad z' = \varphi_{e_1}(e), \quad e_1 \in \mathcal{N}(e_0)$$

(e_0, e_1 otherwise arbitrary) implies that

$$z = \varphi_{e_0}[\varphi_{e_1}^{-1}(z')] \tag{7.7}$$

is analytic at $z' = 0$.[9]

---

[7] In calling $\{\varphi_{e_0}(e)\}$ the "local parameters," we always have in mind variables $z, z', \ldots$ as in $z = \varphi_{e_0}(e)$, $z' = \varphi_{e_1}(e), \ldots$ The distinction between the variables $z, z', \ldots$ and the functions $\varphi_{e_0}(e), \varphi_{e_1}(e), \ldots$ might be emphasized by calling $\varphi_{e_0}(e), \varphi_{e_1}(e), \ldots$ "parametric mappings," reserving the term "local parameters" for the variables $z, z', \ldots$ themselves, but this seems unnecessarily fussy.

[8] To keep the notation simple, we write $f\varphi_{e_0}\varphi_{e_1}^{-1}(z')$ instead of $f\{\varphi_{e_0}[\varphi_{e_1}^{-1}(z')]\}$, and similarly elsewhere.

[9] Then clearly (7.7) is univalent at $z' = 0$ (i.e., in a neighborhood of $z' = 0$).

Remark. In saying that E is an abstract Riemann surface, the qualifying phrase "with respect to a given set of local parameters $\{\varphi_{e_0}(e)\}$ and a corresponding set of parametric neighborhoods $\{\mathcal{N}_{e_0}\}$" is always tacitly understood.

DEFINITION 2. *Given an abstract Riemann surface E, a function $F(e)$ defined on a domain $D \subset E$ is said to be analytic on D if*

$$f(z) = F[\varphi_{e_0}^{-1}(z)]$$

is analytic at $z = 0$ for every $e_0 \in D$.

THEOREM 7.2. *Given an abstract Riemann surface E, every local parameter $z = \varphi_{e_0}(e)$ is analytic in its neighborhood $\mathcal{N}_{e_0}$.*

Proof. The theorem is an immediate consequence of Definitions 1 and 2.

Example 1. Thinking of the points $e \in E$ as complex numbers, let E be any domain in the finite e-plane, equipped with local parameters

$$z = e - e_0, \tag{7.8}$$

where the parametric neighborhood $\mathcal{N}_{e_0}$ is any simply connected domain containing e_0 and contained in E. Then E is an abstract Riemann surface.

Example 2. Let E be any domain in the extended e-plane, equipped with the same local parameters (7.8) if e_0 is finite. However, if $e_0 = \infty$, choose $\mathcal{N}_{e_0}$ to be the exterior of a circle $|z| = R$ whose radius is so large that $\mathcal{N}_{e_0} \subset E$ (think of e_0 as the north pole of the Riemann sphere), and let the corresponding local parameter be $z = 1/e$. Then E is again an abstract Riemann surface.

Equivalence of abstract Riemann surfaces is defined in the natural way:

DEFINITION 3. *Let E and E^* be two abstract Riemann surfaces, equipped with local parameters $\{\varphi_{e_0}(e)\}$ and $\{\psi_{e_0^*}(e^*)\}$, respectively.*[10] *Then E and E^* are said to be equivalent if there exists a homeomorphic mapping $e^* = \Phi(e)$ of E onto E^* which carries every $\varphi_{e_0}(e)$ $[e_0 \in E]$ into a function $\varphi_{e_0}[\Phi^{-1}(e^*)]$ analytic at $e_0^* = \Phi(e_0)$ and every $\psi_{e_0^*}(e^*)$ $[e_0^* \in E^*]$ into a function $\psi_{e_0^*}[\Phi(e)]$ analytic at $e_0 = \Phi^{-1}(e_0^*)$.*

Example 1. Let E be the finite e-plane equipped with local parameters $\varphi_{e_0}(e) = e - e_0$, and let E^* be the finite e^*-plane equipped with local parameters $\psi_{e_0^*}(e^*)$, where $\psi_{e_0^*}(e^*)$ is analytic at e_0^* with power series expansion

$$\psi_{e_0^*}(e^*) = a_1(e^* - e_0^*) + a_2(e^* - e_0^*)^2 + \cdots \qquad (a_1 \neq 0),$$

[10] E and E^* may be identical regarded as topological surfaces, i.e., before choosing the local parameters. This does not imply the equivalence of E and E^*, regarded as abstract Riemann surfaces (see Example 2 below).

so that $\psi_{e_0^*}(e_0^*) = 0$, $\psi'_{e_0^*}(e_0^*) \neq 0$.[11] Then E and E^* are equivalent abstract Riemann surfaces. In fact, we need only choose the homeomorphic mapping in Definition 3 to be $e^* = \Phi(e) = e$.

Example 2. Let E be the finite e-plane equipped with local parameters $\varphi_{e_0}(e) = e - e_0$, and let E^* be the finite e^*-plane equipped with local parameters $\psi_{e_0^*}(e^*) = \chi(e^*) - \chi(e_0^*)$, where[12]

$$\chi(e^*) = \begin{cases} \dfrac{2e^*}{\pi|e^*|} \arctan |e^*| & \text{if } e^* \neq 0, \\ 0 & \text{if } e^* = 0. \end{cases}$$

Then E and E^* are *nonequivalent* abstract Riemann surfaces.[13] In fact, equivalence of E and E^* implies the existence of a homeomorphic mapping $e^* = \Phi(e)$ of E onto E^* carrying $\psi_{e_0^*}(e^*)$ into a function $\psi_{e_0^*}[\Phi(e)]$ analytic at $e_0 = \Phi^{-1}(e_0^*)$ for every $e_0 \in E$. But then $\chi[\Phi(e)]$ must be analytic at every point of the finite e-plane, while

$$\chi[\Phi(e)] \not\equiv \text{const}, \qquad |\chi[\Phi(e)]| < 1.$$

This contradicts Liouville's theorem (Theorem I.16.8), thereby proving that E and E^* are nonequivalent.

THEOREM 7.3. *With the same notation as in Definition* 3, *let E and E^* be equivalent abstract Riemann surfaces, and let $F(e)$ be analytic on a domain $D \subset E$. Then*

$$F^*(e^*) = F[\Phi^{-1}(e^*)]$$

is analytic on the domain $D^* = \Phi(D) \subset E^*$.

Proof. The function

$$z = \varphi_{e_0}\Phi^{-1}\psi_{e_0^*}^{-1}(z')$$

is analytic at $z' = 0$, and moreover $F[\varphi_{e_0}^{-1}(z)]$ is analytic at $z = 0$ for every $e_0 \in D$ (recall Definition 2, p. 223). Therefore

$$F[\varphi_{e_0}^{-1}(z)] = F\varphi_{e_0}^{-1}\varphi_{e_0}\Phi^{-1}\psi_{e_0^*}^{-1}(z') = F\Phi^{-1}\psi_{e_0^*}^{-1}(z') = F^*[\psi_{e_0^*}^{-1}(z')]$$

is analytic at $z' = 0$ for every $e_0^* \in D^*$, being an analytic function of an analytic function. In other words, $F^*(e^*)$ is analytic on D^*, as asserted.

THEOREM 7.4. *Let E and E^* be two domains in the finite complex plane, equipped with local parameters $\varphi_{e_0}(e) = e - e_0$ and $\psi_{e_0^*}(e^*) = e^* - e_0^*$,*

---

[11] The second condition guarantees that $\psi_{e_0^*}(e^*)$ is one-to-one in a neighborhood of e_0^*.

[12] Note that $z = \chi(e^*)$ is a homeomorphic mapping of the finite e^*-plane onto the unit disk $|z| < 1$.

[13] To verify that E^* is an abstract Riemann surface, we observe that

$$z = \psi_{e_0^*}[\psi_{e_1^*}^{-1}(z')] = z' + \chi(e_1^*) - \chi(e_0^*)$$

is an analytic function of z'.

respectively. Then E and E^* are equivalent, regarded as abstract Riemann surfaces, if and only if there exists a (one-to-one) conformal mapping $e^* = \Phi(e)$ of E onto E^*.

Proof. If $\Phi(e)$ exists, then $\Phi(e)$ is analytic by Menshov's theorem (cited on p. 3). But $\varphi_{e_0}(e) = e - e_0$ and $\psi_{e_0^*}(e^*) = e^* - e_0^*$ are obviously analytic (everywhere), and hence $\varphi_{e_0}[\Phi^{-1}(e^*)]$ is analytic at $e_0^* = \Phi(e_0)$ for every $e_0 \in E$ while $\psi_{e_0^*}[\Phi(e)]$ is analytic at $e_0 = \Phi^{-1}(e_0^*)$ for every $e_0^* \in E^*$, i.e., E and E^* are equivalent.

Conversely, suppose E and E^* are equivalent. Then, given any $e_0 \in E$, there is a homeomorphic mapping $e^* = \Phi(e)$ of E onto E^* such that

$$\psi_{e_0^*}[\Phi(e)] = \Phi(e) - e_0^* = \Phi(e) - \Phi(e_0)$$

is analytic at e_0.[14] Therefore $\Phi(e)$ is univalent on E and hence maps E conformally onto E^* (by Theorem II.4.1, Corollary 3).

COROLLARY 1. *Every simply connected domain in the finite plane with a boundary containing more than one point represents the same abstract Riemann surface.*

Proof. Apply Theorem 1.2.

COROLLARY 2. *The class of doubly connected domains in the finite plane contains uncountably many distinct abstract Riemann surfaces.*

Proof. According to Theorem 1.1, an annulus $E: r < |e| < R$ can be mapped conformally onto another annulus $E^*: r^* < |e^*| < R^*$ if and only if

$$\frac{R}{r} = \frac{R^*}{r^*}. \tag{7.9}$$

Therefore E and E^* represent distinct abstract Riemann surfaces if (7.9) fails to hold, as it does for uncountably many distinct values of the ratio on the right, say.

DEFINITION 4. *Let E be an abstract Riemann surface, equipped with local parameters $\{\varphi_{e_0}(e)\}$ and parametric neighborhoods $\{\mathcal{N}_{e_0}\}$, $e_0 \in E$. Then two Jordan curves $\gamma_1 \subset \mathcal{N}_{e_0}$ and $\gamma_2 \subset \mathcal{N}_{e_0}$ with common initial point e_0 are said to form an angle of θ radians with its vertex at e_0 if the Jordan curves $\Gamma_1 = \varphi_{e_0}(\gamma_1)$ and $\Gamma_2 = \varphi_{e_0}(\gamma_2)$, i.e., the images of γ_1 and γ_2 under the mapping $z = \varphi_{e_0}(e)$, form an angle of θ radians with its vertex at $z = 0$.*[15]

---

[14] The power series expansion of $\Phi(e)$ at e_0 is of the form

$$\Phi(e) = \Phi(e_0) + a_1(e - e_0) + a_2(e - e_0)^2 + \cdots \qquad (a_1 \neq 0).$$

[15] In other words, Γ_1 and Γ_2 have tangents τ_1 and τ_2 at $z = 0$, and θ is the angle between τ_1 and τ_2 measured from τ_1 to τ_2 (cf. Sec. I.31).

DEFINITION 5. *Given two abstract Riemann surfaces E and E^*, let $F(e)$ be a one-to-one mapping of a domain $D \subset E$ onto a domain $D^* \subset E^*$. Then $F(e)$ is said to map D conformally onto D^* if $F(e)$ preserves angles (and the directions in which they are measured) at every point $e_0 \in D$.*[16]

THEOREM 7.5. *Let E be an abstract Riemann surface, equipped with local parameters $\{\varphi_{e_0}(e)\}$ and parametric neighborhoods $\{\mathcal{N}_{e_0}\}$, $e_0 \in E$. Then $z = \varphi_{e_0}(e)$ maps $\mathcal{N}_{e_0}$ conformally onto the domain $G_{e_0} = \varphi_{e_0}(\mathcal{N}_{e_0})$.*

Proof. Given any point $e_1 \in \mathcal{N}_{e_0}$ ($e_1 \neq e_0$),[17] let $\gamma_1, \gamma_2 \subset \mathcal{N}_{e_0} \cap \mathcal{N}_{e_1}$ be two Jordan curves forming an angle of θ radians with its vertex at e_1. Then, according to Definition 4, $z' = \varphi_{e_1}(e)$ maps γ_1 and γ_2 into two curves Γ_1' and Γ_2' in the z'-plane forming an angle of θ radians with its vertex at $z' = 0$. On the other hand, $z = \varphi_{e_0}(e)$ maps γ_1 and γ_2 into two Jordan curves Γ_1 and Γ_2 in the z-plane, and we want to show that Γ_1 and Γ_2 also form an angle of θ radians with its vertex at $\zeta = \varphi_{e_0}(e_1) \in G_{e_0}$. But this follows at once from the fact that the mapping $z = \varphi_{e_0}[\varphi_{e_1}^{-1}(z')]$ carrying Γ_1', Γ_2' into Γ_1, Γ_2 is univalent at $z' = 0$ (recall Definition 1, p. 222).

Remark 1. Theorem 7.4 further clarifies the meaning of the requirement that the function (7.7) be analytic at $z' = 0$.

Remark 2. Thus an abstract Riemann surface is not only locally Euclidean in the sense of Definition 3, p. 218, but also "locally conformal" in the sense of Theorem 7.5. Figuratively speaking, an abstract Riemann surface is a topological surface equipped with a "conformal structure," i.e., a set of local parameters $\{\varphi_{e_0}(e)\}$ and a corresponding set of parametric neighborhoods $\{\mathcal{N}_{e_0}\}$ in which they are conformal.

We are now in a position to generalize Theorem 7.4 to the case of arbitrary abstract Riemann surfaces:

THEOREM 7.6. *An abstract Riemann surface E, equipped with local parameters $\{\varphi_{e_0}(e)\}$ and parametric neighborhoods $\{\mathcal{N}_{e_0}\}$, is equivalent to another abstract Riemann surface E^*, equipped with local parameters $\{\psi_{e_0^*}(e^*)\}$ and parametric neighborhoods $\{\mathcal{N}_{e_0^*}\}$, if and only if there exists a conformal mapping $e^* = \Phi(e)$ of E onto E^*.*

---

[16] More precisely, suppose E has local parameters $\{\varphi_{e_0}(e)\}$ and parametric neighborhoods $\{\mathcal{N}_{e_0}\}$, $e_0 \in E$, while E^* has local parameters $\{\psi_{e_0^*}(e^*)\}$ and parametric neighborhoods $\{\mathcal{N}_{e_0^*}\}$, $e_0^* \in E^*$, and let $\gamma_1, \gamma_2 \subset \mathcal{N}_{e_0}$ be two Jordan curves forming an angle of θ radians (measured from γ_1 to γ_2) with its vertex at e_0. Then $\Gamma_1 = F(\gamma_1) \cap \mathcal{N}_{e_0^*}$ and $\Gamma_2 = F(\gamma_2) \cap \mathcal{N}_{e_0^*}$ form an angle of θ radians (measured from Γ_1 to Γ_2) with its vertex at $e_0^* = F(e_0)$, and similarly for every $e_0 \in D$.

[17] If $e_1 = e_0$, there is nothing to prove.

Proof. If $\Phi(e)$ exists, then

$$z = \varphi_{e_0}\Phi^{-1}\psi_{e_0^*}^{-1}(z') \tag{7.10}$$

is a one-to-one conformal mapping of a neighborhood of the point $z' = 0$ onto a neighborhood of the point $z = 0$ (use Theorem 7.5), and hence, by Menshov's theorem (cited on p. 3), the mapping (7.10) is analytic at $z' = 0$. But by Theorem 7.2, $z' = \psi_{e_0^*}(e^*)$ is analytic on $\mathcal{N}_{e_0^*}$, in particular at $e^* = e_0^*$, and hence

$$z = \varphi_{e_0}\Phi^{-1}\psi_{e_0^*}^{-1}\psi_{e_0^*}(e^*) = \varphi_{e_0}[\Phi^{-1}(e^*)]$$

is analytic at e_0^* for every $e_0^* \in E^*$. By the same argument applied to the mappings

$$z' = \psi_{e_0^*}\Phi\varphi_{e_0}^{-1}(z), \qquad z = \varphi_{e_0}(e),$$

we find that

$$z' = \psi_{e_0^*}\Phi\varphi_{e_0}^{-1}\varphi_{e_0}(e) = \psi_{e_0^*}[\Phi(e)]$$

is analytic at e_0 for every $e_0 \in E$. It follows from Definition 3, p. 223 that E and E^* are equivalent.

Conversely, suppose E and E^* are equivalent. Then, given any $e_0 \in E$, there is a homeomorphic mapping $e^* = \Phi(e)$ of E onto E^* such that $\psi_{e_0^*}[\Phi(e)]$ is analytic at e_0. Therefore, since $\varphi_{e_0}^{-1}(z)$ is analytic at $z = 0$, $\psi_{e_0^*}\Phi\varphi_{e_0}^{-1}(z)$ is analytic and hence univalent at $z = 0$. But then, according to Definitions 4 and 5, $\Phi(e)$ maps E conformally onto E^*, and the proof is complete.

Remark. It is clear from Theorem 7.5 that the study of analytic functions defined on abstract Riemann surfaces is confined from the outset to properties of a very general kind, in fact to properties which are invariant under conformal mappings.

33. Triangulations

Given a topological surface E, let $T \subset E$ be a closed Jordan domain which can be mapped homeomorphically onto a triangle Δ in the Euclidean plane. Then T is called a *(topological) triangle*, the points A, B, C mapped into the vertices of Δ are called the *vertices* of T, and the arcs $\widehat{AB}$, $\widehat{BC}$, $\widehat{CA}$ are called the *sides* of T.[18] A triangle T is said to be *oriented* if its vertices A, B and C are assigned a definite order, where arrangements which are cyclic permutations[19] of one another are not distinguished. Thus there are precisely two

---

[18] Of the two arcs joining A to B, $\widehat{AB}$ denotes the one which does not contain C, and similarly for $\widehat{BC}$ and $\widehat{CA}$.

[19] See e.g., G. Birkhoff and S. MacLane, *op. cit.*, p. 133.

orientations of T, one corresponding to the arrangement ABC (equivalently BCA or CAB) and the other to ACB (equivalently CBA or BAC).

DEFINITION 1. *Given a two-dimensional manifold E, suppose E is the union of a family of triangles $\mathscr{T}$ such that*

1. *If e is an interior point of a triangle $T \in \mathscr{T}$, then e belongs to no other triangle $T' \in \mathscr{T}$;*
2. *If e is a boundary point but not a vertex of a triangle $T \in \mathscr{T}$, then e also belongs to a unique triangle $T' \in \mathscr{T}$ distinct from T, which shares a side but no other points with T;*
3. *Two triangles in $\mathscr{T}$ with no side in common either share a unique vertex or else are disjoint;*
4. *Every point $e \in E$ has a neighborhood $\mathscr{N}(e)$ covered by a finite number of triangles in $\mathscr{T}$.*

Then E is said to be triangulable, and $\mathscr{T}$ is called a triangulation of E.

Remark. It can be shown that a two-dimensional manifold is triangulable if and only if it has a countable base.[20] Therefore every topological surface is triangulable (recall Definition 4, p. 218). However, the existence of triangulations for all the surfaces of interest here is easily established directly.

THEOREM 7.7. *Let $\mathscr{T}$ be a triangulation of a topological surface E, and let A be a vertex of any triangle $T \in \mathscr{T}$. Then $\mathscr{T}$ contains a finite number of triangles $T = T_1, \ldots, T_n$ $(n \geqslant 3)$ forming a star of triangles at A in the following sense;*

1. *A is a vertex of every triangle $T_1, \ldots, T_n$;*
2. *For every $k = 1, \ldots, n$, the triangles T_k and T_{k+1} share a side $(T_{n+1} = T_1)$;*
3. *$T_1, \ldots, T_n$ cover a neighborhood of A.*[21]

Proof. According to properties 1 and 2 of a triangulation, A cannot be an interior point or a boundary point other than a vertex of any other triangle in $\mathscr{T}$. Moreover, by property 4, A has a neighborhood $\mathscr{N} = \mathscr{N}(A)$ covered by a finite number of triangles

$$T^{(1)}, \ldots, T^{(m)}, \tag{7.11}$$

and no other triangle T^* can intersect $\mathscr{N}$, since otherwise one or several of the triangles (7.11) would contain interior points of T^*, contrary to property 1. In particular, the original triangle T must coincide with one of the triangles (7.11).

Now let AB be one of the two sides of T emanating from A. Then, by property 2, there exists a triangle T' distinct from T which has AB as a

[20] See Problem 7.9 and pp. 73–75 of the book by Stoïlow (cited in the Bibliography on pp. 351–353). The so-called *Prüfer manifold* (ibid., pp. 72–75) does not have a countable base, and hence is nontriangulable.

[21] I.e., there is a neighborhood $\mathscr{N}(A)$ contained in $T_1 \cup \cdots \cup T_n$.

side. Since T' has points in every neighborhood of A, T' must coincide with one of the triangles (7.11). Let AB' be the side of T' which emanates from A but is distinct from the side AB shared by T and T'. By property 2, AB' cannot belong to T, and moreover, by the same property, there exists a triangle T'' distinct from T and T' which has AB' as a side. As before, T'' must coincide with one of the triangles (7.11). Continuing this construction, we eventually obtain a total of n distinct triangles

$$T = T_1, \quad T' = T_2, \quad T'' = T_3, \ldots, T_n, \tag{7.12}$$

where T_k and T_{k+1} ($k = 1, \ldots, n-1$) each share a side emanating from A. Since $\mathcal{N}$ is covered by m triangles (7.11) in all, we must have $3 \leqslant n \leqslant m$. Moreover, T_n and T_1 must share a side (which is neither the common side of T_{n-1} and T_n nor the original side AB), since otherwise the number of triangles (7.12) would exceed n. Mapping some neighborhood of A (either $\mathcal{N}$ or a subset of $\mathcal{N}$) homeomorphically onto a simply connected domain in the z-plane, we see that images of $T_1 \cap \mathcal{N}, \ldots, T_n \cap \mathcal{N}$ cover a neighborhood of z_0, forming a star of triangles at z_0 in the sense described in the statement of the theorem. Therefore the triangles (7.12) cover a corresponding neighborhood of A, forming a star at A, and the proof is complete.

DEFINITION 2. *Given any two triangles T' and T'' belonging to a triangulation $\mathcal{T}$ of a topological surface E, let*

$$T' = T_1, \quad T_2, \ldots, T_n = T'' \tag{7.13}$$

be a finite set of triangles in $\mathcal{T}$ such that each T_k shares a side with T_{k+1} ($k = 1, \ldots, n-1$). Then (7.13) is called a chain of triangles joining T' to T'', and the triangles T' and T'' are said to be chained (to each other).[22]

THEOREM 7.8. *Any two triangles T' and T'' belonging to a triangulation $\mathcal{T}$ of a topological surface E are chained.*

Proof. Let $\mathcal{T}_1 \subset \mathcal{T}$ be the set of all triangles chained to T',[23] and suppose $\mathcal{T}_1 \neq \mathcal{T}$. Then the set $\mathcal{T}_2 = \mathcal{T} - \mathcal{T}_1$, consisting of all triangles not chained to T' is nonempty, and so are the sets

$$E_1 = \bigcup_{T \in \mathcal{T}_1} T, \qquad E_2 = \bigcup_{T \in \mathcal{T}_2} T.$$

Clearly $E_1 \cap E_2 = 0$ and $E = E_1 \cup E_2$. If $e \in E_1$, there are only three possibilities:

1. e_1 is an interior point of a triangle $T \in \mathcal{T}_1$;
2. e_1 is a boundary point but not a vertex of a triangle $T \in \mathcal{T}_1$, and hence belongs to a common side of two triangles in $\mathcal{T}$;
3. There is a star of triangles of $\mathcal{T}_1$ at e_1.

---

[22] In particular, every triangle is regarded as chained to itself.

[23] Thus $\mathcal{T}_1$ contains T' itself, the three triangles sharing sides with T', and so on.

In each case, every triangle containing e_1 belongs to $\mathscr{T}_1$. Therefore E_1 is open, and the same argument shows that E_2 is open. Thus E is the union of two nonempty disjoint sets E_1 and E_2, each containing no limit points of the other. But this is impossible, since E is connected, by hypothesis. Therefore $\mathscr{T}_1 = \mathscr{T}$, and hence $T'' \in \mathscr{T}_1$, i.e., T'' is chained to T', as asserted.

THEOREM 7.9. *A triangulation $\mathscr{T}$ of a topological surface E can contain only countably many triangles.*

Proof. Starting from any triangle $T' \in \mathscr{T}$, we first count T', then the three triangles sharing sides with T', then the triangles sharing sides with these triangles, and so on. According to Theorem 7.8, we eventually count any other triangle $T'' \in \mathscr{T}$, i.e., $\mathscr{T}$ is countable, as asserted.

THEOREM 7.10. *A triangulation $\mathscr{T}$ of a topological surface E contains finitely many triangles if and only if E is closed.*

Proof. If $\mathscr{T}$ is finite, then E is covered by a finite number of triangles. But every triangle is compact, being the homeomorphic image of a closed (Jordan) domain, and hence E is compact, being the union of a finite number of compact sets. It follows that every infinite subset of E has a limit point in E (recall Problem I.3.23), i.e., E is closed in the sense of Definition 5, p. 218.

Conversely, suppose E is closed but has an infinite triangulation $\mathscr{T}$. Then, choosing one interior point in each triangle $T \in \mathscr{T}$, we obtain an infinite set $F \subset E$ which has no limit point in E. In fact, every point $e \in E$ has a neighborhood $\mathscr{N}(e)$ covered by a finite number of triangles in $\mathscr{T}$, and hence $\mathscr{N}(e)$ can contain only finitely many points of F. This contradicts the assumption that E is closed, thereby showing that $\mathscr{T}$ is finite.

Example 1. Figure 7.3 shows a triangulation of the sphere (after stereographic projection), while Figure 7.4 shows a triangulation of the torus,

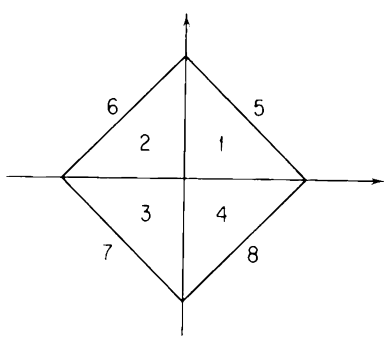


FIGURE 7.3

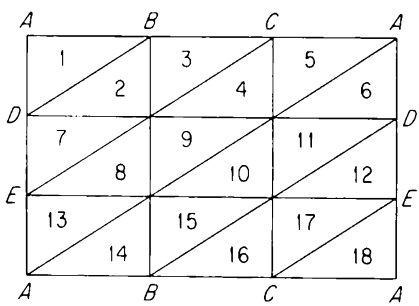


FIGURE 7.4

represented by a rectangle with pairs of points on opposite sides identified ("pasted together") as in Example 4, p. 220. In both cases, the triangulations are finite (8 triangles for the sphere and 16 for the torus). This is to be expected, since both surfaces are closed.

Example 2. Figure 7.5 shows a triangulation of the Möbius strip, represented by a rectangle with pairs of points on the short ends identified if they are symmetric with respect to the origin (just as in Example 5, p. 220). The triangulation is now infinite, since the Möbius strip is an open surface.

DEFINITION 3. *Given a triangulation $\mathscr{T}$ of a topological surface E, let T' and T'' be two oriented triangles in $\mathscr{T}$ sharing a side with end points A and B. Then T' and T'' are said to be coherently oriented if A precedes B in the orientation chosen for one triangle, while B precedes A in the orientation chosen for the other triangle.*

Remark. Thus, figuratively speaking, if the common side of two coherently oriented triangles T' and T'' is traversed in one direction regarded as a side of T'', it is traversed in the opposite direction regarded as a side of T''.

Example. The triangles T_1 and T_6 shown in Figure 7.5 are coherently oriented if T_6 has the orientation corresponding to the dotted arrows, but not if T_6 has the orientation corresponding to the solid arrows.

DEFINITION 4. *Given a triangulation $\mathscr{T}$ of a topological surface E, let $\mathscr{C} \subset \mathscr{T}$ be a chain of triangles $T_1, \ldots, T_n$ joining T_1 to T_n. Then $\mathscr{C}$ is said to be coherently oriented if T_k and T_{k+1} are coherently oriented for every $k = 1, \ldots, n-1$.*

DEFINITION 5. *Given a triangulation $\mathscr{T}$ of a topological surface E, let $\mathscr{C} \subset \mathscr{T}$ be a coherently oriented chain joining two triangles $T', T'' \in \mathscr{T}$.*[24] *Then E is said to be orientable (or two-sided) if the orientation of T'' depends only on that of T' and not on the choice of $\mathscr{C}$, for every $T', T'' \in \mathscr{T}$. Otherwise, E is said to be nonorientable (or one-sided).*

Remark. Definition 5 tacitly relies upon the fact that the orientability of E is independent of the choice of $\mathscr{T}$.[25]

Example 1. The plane, sphere and torus are orientable surfaces (see Problem 7.12).

Example 2. The Möbius strip is one-sided, as can be seen by considering two coherently oriented chains joining T_1 to T_6 in Figure 7.5, one consisting

---

[24] The existence of $\mathscr{C}$ for arbitrary $T', T'' \in \mathscr{T}$ is guaranteed by Theorem 7.8.

[25] See e.g., P. S. Aleksandrov, *Combinatorial Topology* (translated by H. Komm), Graylock Press, Rochester, N.Y., (1956–7), Vol. I, p. 95, and Vol. II, p. 156.

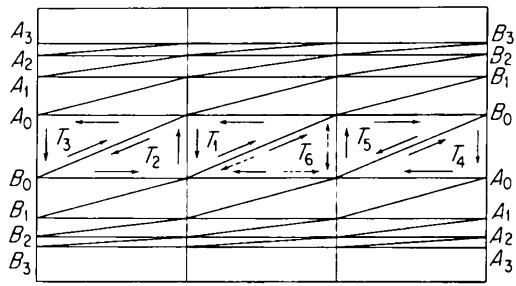


FIGURE 7.5

of just T_1 and T_6 (where T_6 has the orientation corresponding to the dotted arrows), the other consisting of all six triangles $T_1, \ldots, T_6$. In fact, starting from T_1 and using the chain $T_1, \ldots, T_6$, we arrive at T_6 with the orientation corresponding to the solid arrows.

34. Interior Mappings

DEFINITION. *Given two topological surfaces E and E^*, let $e^* = f(e)$ be a continuous mapping of E into E^*.*[26] *Suppose E has a triangulation $\mathcal{T}$ such that, given any triangle $T \in \mathcal{T}$,*

1. *$f(e)$ is homeomorphic on* int (T);[27]
2. *$f(e)$ is homeomorphic in a neighborhood of every boundary point of T except possibly vertices of T.*

Then $f(e)$ is called an interior mapping (in the sense of Stoïlow).

Remark 1. Obviously, every homeomorphism of E into E^* is an interior mapping.

Remark 2. It can be shown that the following are necessary and sufficient conditions for a continuous mapping $e^* = f(e)$ of E into E^* to be an interior mapping:

1. $f(e)$ carries open sets on E into open sets on E^*;
2. If $f(e)$ carries a continuum $C \subset E$ into a single point of E^*, then C itself consists of a single point.

The necessity is left as an exercise for the reader, but the sufficiency is considerably harder.[28]

We now study the case where the surface E^* is the extended complex

---

[26] The choice of the word "into" instead of "onto" is important here (cf. p. I.26).
[27] By int (T), we mean the set of all interior points of T (cf. Problem I.4.14).
[28] See S. Stoïlow, *op. cit.*, Chap. 5, Sec. 2.

plane, i.e., the (Riemann) sphere Σ. The points of E^* are then complex numbers (including ∞), and we write w instead of e^*.[29]

THEOREM 7.11. *If there exists an interior mapping $w = f(e)$ of a topological surface E into the sphere Σ, then E is orientable.*

Proof. Let $\mathscr{T}$ be the triangulation figuring in the definition of $w = f(e)$, and let T be any triangle in $\mathscr{T}$. Then $f(e)$ maps T into a triangle Δ on the sphere Σ, and the orientation of T determines a definite orientation of Δ. Suppose E is nonorientable. Then there are two triangles T', $T'' \in \mathscr{T}$ and two coherently oriented chains $\mathscr{C}_1$ and $\mathscr{C}_2$ joining T' to T'' such that T' has the same orientation in both $\mathscr{C}_1$ and $\mathscr{C}_2$, while T'' has one orientation in $\mathscr{C}_1$ and the opposite orientation in $\mathscr{C}_2$. Therefore the same is true of the triangles $\Delta' = f(T')$, $\Delta'' = f(T'')$ and the chains $\mathscr{C}_1^* = f(\mathscr{C}_1)$, $\mathscr{C}_1^* = f(\mathscr{C}_2)$. But this is impossible, since the sphere is an orientable surface. Hence E must be orientable, as asserted.

The converse of Theorem 7.11 will be established, after proving a preliminary result on the extension of homeomorphisms:

LEMMA. *Given two topological surfaces E and E^*, let $D \subset E$ and $D^* \subset E^*$ be closed Jordan domains with boundaries Γ and Γ^*, respectively. Suppose $e^* = f(e)$ maps Γ homeomorphically onto Γ^*. Then there exists a homeomorphism $e^* = F(e)$ which maps D onto D^* and coincides with $f(e)$ on Γ.*[30]

Proof. Let $z = \Phi(e)$ and $z^* = \Psi(e^*)$ be homeomorphic mappings of D and D^* onto the unit disks $K: |z| < 1$ and $K^*: |z^*| < 1$, with boundaries C and C^* (why do such mappings exist?). Then $\tilde{f}(z) = \Psi f \Phi^{-1}(z)$ is a homeomorphism of C onto C^*. If $\tilde{f}(z)$ carries the point $e^{i\theta} \in C$ into the point $e^{i\theta^*} \in C^*$, where $\theta^* = \lambda(\theta)$, then $\lambda(\theta)$ is one-to-one in the interval $[0, 2\pi)$. It follows that $\tilde{F}(z) = re^{i\lambda(\theta)}$ $(0 \leqslant r \leqslant 1)$ is a homeomorphism of K onto K^*, which coincides with $\tilde{f}(z)$ on C. Therefore $F(e) = \Psi^{-1}\tilde{F}\Phi(e)$ is a homeomorphism of D onto D^*, which coincides with

$$\Psi^{-1}\tilde{f}\Phi(e) = \Psi^{-1}\Psi f \Phi^{-1}\Phi(e) = f(e)$$

on Γ, as required.

THEOREM 7.12. *If E is an orientable topological surface, there exists an interior mapping $w = f(e)$ of E into the sphere Σ.*

---

[29] In choosing the symbol w instead of z, we anticipate the case where E itself is a domain in the extended complex plane (see Sec. 36). The points of E will then be denoted by z.

[30] I.e., $F(e) = f(e)$ for every $e \in \Gamma$, and by the same token, $F^{-1}(e^*) = f^{-1}(e^*)$ for every $e^* \in \Gamma^*$.

Proof. Given a triangulation $\mathscr{T}$ of E, we divide every triangle $T \in \mathscr{T}$ into six smaller triangles by drawing the medians of T, as shown in Figure 7.6.[31] This gives rise to a new triangulation $\mathscr{T}'$, called the *barycentric subdivision* of the original triangulation $\mathscr{T}$. Given any triangle $T \in \mathscr{T}$, let the vertices of T be labelled a, let the midpoints of the sides of T be labelled b, and let the point of intersection of the medians of T (i.e., the centroid of T) be labelled c. Then every triangle in $\mathscr{T}'$ has vertices a, b and c, as shown in the figure. Since E is orientable, once an orientation has been chosen for the triangles in $\mathscr{T}$ (cf. Problem 7.11), we can divide the triangles in $\mathscr{T}'$ into two classes, one consisting of shaded triangles, the other consisting of unshaded triangles, where the choice is made as follows: Given any triangle $T' \in \mathscr{T}'$, one side ab of T' is contained in a side δ of some triangle $T \in \mathscr{T}$, and the orientation of T assigns δ an orientation, specifically an initial point and a final point. If ab contains the initial point of δ, we shade T', while if ab contains the final point of δ, we leave T' unshaded. In this way, two triangles of the same class never share a side (cf. Figure 7.6).[32]

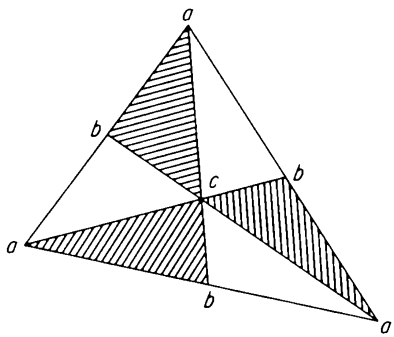


FIGURE 7.6

Next, choosing three distinct points α, β and γ on the equator of the sphere Σ, we map the side ab of every triangle $T' \in \mathscr{T}'$ homeomorphically onto the arc $\widehat{\alpha\beta}$ ($\gamma \notin \widehat{\alpha\beta}$). Similarly, we map the sides bc and ca onto the arcs $\widehat{\beta\gamma}$ and $\widehat{\gamma\alpha}$. Then, using the lemma, we extend this mapping of the sides to a homeomorphic mapping of every shaded triangle onto the upper hemisphere and of every unshaded triangle onto the lower hemisphere. The result is a continuous mapping $w = f(e)$ of E into Σ (in fact, onto Σ), which is one-to-one on every triangle $T' \in \mathscr{T}'$ and also in a neighborhood of every boundary point of T' except the vertices a, b and c. In other words, $w = f(e)$ is an interior mapping of E into Σ, and the proof is complete.

The behavior of an interior mapping $w = f(e)$ near the vertices of the underlying triangulation $\mathscr{T}$ is of key importance. Suppose $f(e)$ fails to be homeomorphic in every neighborhood of a point $e_0 \in E$. Then e_0 must be

---

[31] More exactly, we draw the medians of the Euclidean triangle Δ homeomorphic to T, and then transform back from Δ to T.

[32] Note that if E is nonorientable, this last assertion breaks down, as can be seen by making the barycentric subdivision of the triangulation of the Möbius strip given in Figure 7.5.

a vertex of a triangle $T \in \mathscr{T}$. In fact, according to Theorem 7.7, there is a star of triangles of $\mathscr{T}$ at e_0. Since $\mathscr{T}$ is countable (by Theorem 7.9), so is the set V of all such exceptional points e_0. Moreover, V can have no limit points on E (why?).

More information about the behavior of $f(e)$ at e_0 is given by

THEOREM 7.13. *Let E be a topological surface with triangulation $\mathscr{T}$, which has an interior mapping $w = f(e)$ into the sphere Σ. Suppose $f(e)$ fails to be homeomorphic in every neighborhood of a point $e_0 \in E$, and let $T_1, \ldots, T_n$ be the star of triangles of $\mathscr{T}$ at e_0. Then there is a deleted neighborhood $\mathscr{N}'(w_0)$, where $w_0 = f(e_0)$, every point of which has precisely m distinct inverse images in $T_1 \cup \cdots \cup T_n$, where $m \leqslant n$.*

Proof. Every neighborhood of e_0 contains distinct points at which $f(e)$ takes the same value, but a given value cannot be taken more than m times, where $m \leqslant n$, since $f(e)$ is homeomorphic on the interior of every triangle T_k $(k = 1, \ldots, n)$ and also in a neighborhood of every boundary point of T_k except the point e_0 itself.[33] Let $\mathscr{N}(w_0)$ be any neighborhood of $w_0 = f(e_0)$ which does not contain the images of the sides of $T_1, \ldots, T_n$ lying opposite e_0, so that $\mathscr{N}(w_0)$ is covered by the images of $T_1, \ldots, T_n$, and let $\mathscr{N}'(w_0)$ be the deleted neighborhood $\mathscr{N}(w_0) - \{w_0\}$. Let m be the maximum number of distinct inverse images in $T_1 \cup \cdots \cup T_n$ belonging to any point of $\mathscr{N}'(w_0)$, and let w' be a point of $\mathscr{N}'(w_0)$ which in fact has m distinct inverse images $e_1', \ldots, e_m'$ in $T_1 \cup \cdots \cup T_n$.[34] Then $T_1 \cup \cdots \cup T_n$ contains disjoint neighborhoods $\mathscr{N}(e_1'), \ldots, \mathscr{N}(e_m')$ in which $f(e)$ is homeomorphic. Therefore each of the sets $f[\mathscr{N}(e_1')], \ldots, f[\mathscr{N}(e_m')]$ is a neighborhood, and their union contains a neighborhood $\mathscr{N}(w')$. But then every point $w \in \mathscr{N}(w')$ has precisely m inverse images in $T_1 \cup \cdots \cup T_n$, i.e., one in each neighborhood $\mathscr{N}(e_j')$, and hence the set M of points in $\mathscr{N}'(w_0)$ which have precisely m distinct inverse images in $T_1 \cup \cdots \cup T_n$ is open.

Now let $w'' \in \mathscr{N}'(w_0)$ be a limit point of M, and suppose w'' has fewer than m distinct inverse images in $T_1 \cup \cdots \cup T_n$ which we denote by $e_1'', \ldots, e_\mu''$ $(\mu < m)$.[35] Then, as before, $T_1 \cup \cdots \cup T_n$ contains disjoint neighborhoods $\mathscr{N}(e_1''), \ldots, \mathscr{N}(e_\mu'')$ in which $f(e)$ is homeomorphic. Clearly, every neighborhood of w'' contains points with inverse images in $T_1 \cup \cdots \cup T_n$ but not in $\mathscr{N}(e_1'') \cup \cdots \cup \mathscr{N}(e_\mu'')$. Hence there are sequences $\{w_j\}$ and $\{e_j\}$, where $w_j = f(e_j)$, such that $\{w_j\}$ approaches

---

[33] Note that if $f(e)$ takes the same value η in the interiors of two adjacent triangles T_k and T_{k+1}, then $f(e)$ cannot take the value η at a common boundary point of T_k and T_{k+1} (why is this true?).

[34] Of course, a point of $\mathscr{N}'(w_0)$ may have other inverse images which do not belong to $T_1 \cup \cdots \cup T_n$.

[35] By the very definition of m, w'' cannot have more than m distinct inverse images in $T_1 \cup \cdots \cup T_n$.

w'' while $\{e_j\}$ is bounded (being contained in $T_1 \cup \cdots \cup T_n$). It follows that there are subsequences $\{w_{j_k}\}$ and $\{e_{j_k}\}$, where $w_{j_k} = f(e_{j_k})$, such that $\{w_{j_k}\}$ approaches w'' while $\{e_{j_k}\}$ approaches a point e''. Since

$$f(e'') = \lim_{k\to\infty} f(e_{j_k}) = \lim_{k\to\infty} w_{j_k} = w'' \in \mathcal{N}'(w_0),$$

e'' must be an inverse image of w'' in $T_1 \cup \cdots \cup T_n$ distinct from $e_1'', \ldots, e_\mu''$, contrary to assumption. This contradiction shows that w'' has precisely m distinct inverse images in $T_1 \cup \cdots \cup T_n$. In other words, every limit point of M belonging to $\mathcal{N}'(w_0)$ also belongs to M. But then, since M is open, it follows from Theorem I.4.16 that M coincides with $\mathcal{N}'(w_0)$, and the theorem is proved.

COROLLARY 1. *The triangles* $\Delta_1 = f(T_1), \ldots, \Delta_n = f(T_n)$ *cover the* neighborhood $\mathcal{N}(w_0)$ *m times in the following sense*:

1. w_0 *is a vertex of every triangle* $\Delta_1, \ldots, \Delta_n$;
2. *For every* $k = 1, \ldots, n$, *the triangles* Δ_k *and* Δ_{k+1} *share a side* $(\Delta_{n+1} = \Delta_1)$;
3. *Every point* $w \in \mathcal{N}'(w_0)$ *belongs to m distinct triangles.*[36]

COROLLARY 2. *The mapping*

$$z = (w - w_0)^{1/m} = [f(e) - f(e_0)]^{1/m} \tag{7.14}$$

is homeomorphic in a neighborhood $\mathcal{N}_{e_0}$ *of the point* $e = e_0$ *if the single-valued branches of* (7.14) *are chosen consistently in the following sense*: *For every* $k = 1, \ldots, n$, *the branch chosen in* Δ_k *coincides with the branch chosen in* Δ_{k+1} *on the common side of* Δ_k *and* Δ_{k+1} $(\Delta_{n+1} = \Delta_1)$.

Proof. The prescribed choice of branches transforms the triangles $\Delta_1, \ldots, \Delta_n$, covering a neighborhood of $w = w_0$ m times, into a set of topological triangles $\delta_1, \ldots, \delta_n$ in the z-plane (each with a vertex at $z = 0$) covering a neighborhood K of $z = 0$ just once in the sense of Corollary 1, i.e., forming a star of triangles at $z = 0$ in the sense of Theorem 7.7. The mapping $z = [f(e) - f(e_0)]^{1/m}$ is single-valued by construction, and hence distinct values of z have distinct inverse images $e \in E$. Moreover, there is just one inverse image in $T_1 \cup \cdots \cup T_n$ corresponding to a given point in K. In fact, suppose $z' \in K$ $(z' \neq 0)$ has at least two distinct inverse images e' and e''. Since z' is one of the m values of $(w' - w_0)^{1/m}$ and the set of inverse images of the remaining $m - 1$ values contains at least $m - 1$ distinct points in $T_1 \cup \cdots \cup T_n$ (none of which coincides with e' or e''), we find that the m values of $(w' - w_0)^{1/m}$ have at least $m + 1$ distinct inverse images in $T_1 \cup \cdots \cup T_n$.

---

[36] If w lies on the sides of some of the triangles $\Delta_1, \ldots, \Delta_n$, then in saying that w belongs to m triangles, every pair of triangles with w as a common boundary point is counted as one triangle. Note that if $m = 1$, then $\Delta_1, \ldots, \Delta_n$ form a star of triangles at w_0.

But the latter are also inverse images of the point $w' \neq 0$ under the mapping $w = f(e)$, and the existence of these inverse images contradicts Theorem 7.13. Therefore (7.14) is one-to-one in a neighborhood of the point $t = 0$,[37] and hence also in a neighborhood $\mathcal{N}_{e_0}$ of the point $e = e_0$. Since the mapping (7.14) is continuous on $\mathcal{N}_{e_0}$ by construction, it follows from Theorem I.6.1 that (7.14) is homeomorphic on $\mathcal{N}_{e_0}$, as asserted.

Remark 1. In Corollary 2, it is tacitly assumed that $w_0 = f(e_0)$ is finite. If $w_0 = \infty$, the corollary remains true if we replace $f(e) - f(e_0)$ by $1/f(e_0)$ and (7.14) by

$$z = w^{1/m} = [f(e)]^{-1/m}.$$

Remark 2. There are only countably many points where the integer m figuring in Theorem 7.13 and its corollaries can exceed 1, and these points cannot have a limit point on E.

35. Riemann Covering Surfaces

THEOREM 7.14. *Let E be a topological surface which has an interior mapping $w = f(e)$ into the sphere Σ. For every $e_0 \in E$, let E be equipped with the local parameter*

$$z = [f(e) - f(e_0)]^{1/m} \tag{7.15}$$

and the parametric neighborhood $\mathcal{N}_{e_0}$, where m and $\mathcal{N}_{e_0}$ are the same as in Theorem 7.13 and its corollaries, and (7.15) is replaced by

$$z = [f(e)]^{-1/m}$$

if $f(e_0) = \infty$.[38] *Then E is an abstract Riemann surface.*

Proof. If $e' \in \mathcal{N}_{e_0}$ and if $z = [f(e) - f(e')]^{1/m'}$ is the corresponding local parameter, then

$$z = [f(e') - f(e_0) + \zeta^{m'}]^{1/m},$$

and we need only verify that z is analytic at $\zeta = 0$ (recall Definition 1, p. 222). This is obvious if $m = 1$, and follows from the fact that $f(e') \neq f(e_0)$ if $m > 1$ [otherwise $f(e)$ fails to be homeomorphic on $\mathcal{N}_{e_0}$, contrary to Corollary 2].

THEOREM 7.15. *Any orientable topological surface can be equipped with a conformal structure making it into an abstract Riemann surface.*

Proof. Combine Theorems 7.12 and 7.14.

---

[37] The point $z = 0$ itself obviously has the unique inverse image $e = e_0$ in $T_1 \cup \cdots \cup T_n$.

[38] If $f(e)$ is homeomorphic in a neighborhood of e_0, we automatically set $m = 1$.

DEFINITION 1. *By a Riemann (covering) surface is meant a topological surface E (necessarily orientable), equipped with an interior mapping $w = f(e)$ of E into the sphere Σ.*

Remark. The phrase "Riemann surface," without further qualification, will always mean a Riemann covering surface. Correspondingly, the adjective "abstract" will never be omitted in the phrase "abstract Riemann surface."

DEFINITION 2.[39] *Let E be a Riemann covering surface, and let $w = f(e)$ be the corresponding interior mapping (of E into the sphere Σ). Then a function $F(e)$ is said to be analytic at a point $e_0 \in E$ if $F[\varphi_{e_0}^{-1}(z)]$ is analytic at $z = 0$, where*

$$z = \varphi_{e_0}(e) = \begin{cases} [f(e) - f(e_0)]^{1/m} & \text{if } f(e_0) \neq \infty, \\ [f(e)]^{-1/m} & \text{if } f(e_0) = \infty. \end{cases}$$

Moreover, $F(e)$ is said to have a pole of order α at e_0 if $F[\varphi_{e_0}^{-1}(z)]$ has a pole of order α at $z = 0$.

THEOREM 7.16. *Given a Riemann covering surface E, let $\mathcal{T}$ and $w = f(e)$ be the corresponding triangulation and interior mapping. Then $f(e)$ is analytic at every point $e_0 \in E$ where $f(e)$ is finite, while $f(e)$ has a pole of order m at every point e_0 where $f(e)$ is infinite and the images $f(T_1), \ldots, f(T_n)$ of the star of triangles $T_1, \ldots, T_n \in \mathcal{T}$ at e_0 cover a neighborhood of $w_0 = f(e_0)$ m times.*

Proof. If $f(e_0) \neq \infty$, then

$$F[\varphi_{e_0}^{-1}(z)] = f(e_0) + z^m,$$

which is obviously analytic at $z = 0$, while if $f(e_0) = \infty$, we have

$$F[\varphi_{e_0}^{-1}(z)] = z^{-m}.$$

Remark. Thus, in the sense of Definition 2, the interior mapping $w = f(e)$ is meromorphic on E, i.e., analytic everywhere on E except for poles. Note that there can be no more than countably many poles, with no limit points on E.

The notion of equivalent Riemann covering surfaces is introduced in the natural way:

DEFINITION 3. *Let E and E^* be two Riemann covering surfaces, and let the corresponding interior mappings be $w = f(e)$ and $w = f^*(e^*)$, respectively. Then E and E^* are said to be equivalent if there exists a homeomorphic mapping $e^* = \Phi(e)$ of E onto E^* carrying $f(e)$ into $f^*(e^*)$ and conversely, i.e., if*

$$f^*[\Phi(e)] = f(e), \qquad f[\Phi^{-1}(e^*)] = f^*(e^*).$$

---

[39] Cf. Definition 2, p. 223.

THEOREM 7.17. *Two equivalent Riemann covering surfaces E and E^* represent equivalent abstract Riemann surfaces when equipped with conformal structures as prescribed by Theorem* 7.14.

Proof. According to Definition 3, p. 223, it must be shown that $\psi_{e_0^*}\Phi\varphi_{e_0}^{-1}(z)$ is analytic at $z = 0$ while $\varphi_{e_0}\Phi^{-1}\psi_{e_0^*}^{-1}(z')$ is analytic at $z' = 0$, where

$$z = \varphi_{e_0}(e) = [f(e) - f(e_0)]^{1/m},$$
$$z' = \psi_{e_0^*}(e^*) = [f^*(e^*) - f^*(e_0^*)]^{1/m}$$

if $f(e_0) \neq \infty$, while

$$z = \varphi_{e_0}(e) = [f(e)]^{-1/m},$$
$$z' = \psi_{e_0^*}(e^*) = [f^*(e^*)]^{-1/m}$$

if $f(e_0) = \infty$. In either case, we have

$$\psi_{e_0^*}\Phi\varphi_{e_0}^{-1}(z) = z, \qquad \varphi_{e_0}\Phi^{-1}\psi_{e_0^*}^{-1}(z') = z',$$

and the theorem is proved.

The converse of Theorem 7.17 is not true, i.e., two equivalent abstract Riemann surfaces may represent different Riemann covering surfaces when equipped with suitable interior mappings $w = f(e)$ and $w = f^*(e^*)$, as shown by the following

Example. Thinking of e and e^* as complex numbers, let E and E^* both be the extended complex plane (Riemann sphere), and let the corresponding interior mappings be $w = f(e) = e$ and $w = f^*(e^*) = e^{*2}$. Then $f(e)$ is trivially a homeomorphic mapping of E onto the sphere Σ, while $f^*(e^*)$ is homeomorphic on every triangle of the triangulation $\mathscr{T}$ of Σ shown in Figure 7.3, p. 230, and also in a neighborhood of every boundary point of any triangle $T \in \mathscr{T}$ except at the two vertices $e^* = 0$ and $e^* = \infty$. Therefore $w = f(e)$ and $w = f^*(e^*)$ are both interior mappings, the first making E into an abstract Riemann surface with local parameters

$$z = \varphi_{e_0}(e) = \begin{cases} e - e_0 & \text{if } e_0 \neq \infty, \\ \dfrac{1}{e} & \text{if } e_0 = \infty, \end{cases}$$

the second making E^* into an abstract Riemann surface with local parameters

$$z' = \psi_{e_0^*}(e^*) = \begin{cases} e^{*2} - e_0^{*2} & \text{if } e_0^* \neq 0, \infty, \\ e^* & \text{if } e_0^* = 0, \\ \dfrac{1}{e^*} & \text{if } e_0^* = \infty. \end{cases}$$

Regarded as abstract Riemann surfaces, E and E^* are equivalent, as we see at once by choosing $e^* = \Phi(e) = e$ as the homeomorphic mapping of E onto E^*. In fact,

$$\psi_{e_0^*}\Phi\varphi_{e_0}^{-1}(z) = \begin{cases} 2e_0^* z + z^2 & \text{if } e_0^* \neq 0, \infty, \\ z & \text{if } e_0^* = 0, \infty, \end{cases}$$

while

$$\varphi_{e_0}\Phi^{-1}\psi_{e_0^*}^{-1}(z') = \begin{cases} (e_0^2 + z')^{1/2} - e_0 & \text{if } e_0 \neq 0, \infty, \\ z' & \text{if } e_0 = 0, \infty. \end{cases}$$

On the other hand, E and E^* are not equivalent, regarded as Riemann covering surfaces. In fact, since $w = f(e)$ is one-to-one on E while $w = f^*(e^*)$ is not one-to-one on E^*, there can be no homeomorphic mapping of E onto E^* carrying $f(e)$ into $f^*(e^*)$.

Let E be a Riemann covering surface, and let $\mathscr{T} = \{T_k\}$ and $w = f(e)$ be the corresponding triangulation and interior mapping into the sphere Σ.[40] Given any $T_k \in \mathscr{T}$, let $\Delta_k = f(T_k)$, and then imagine that every pair of triangles Δ_k, Δ_l $(k \neq l)$ sharing a side is "pasted together" along the common side. This construction leads to countably many "wreathes" of triangles, each forming an m-fold covering (in the sense of Theorem 7.13, Corollary 2) of a neighborhood $\mathscr{N}(z_0)$, where $w_0 = f(e_0)$ is a common vertex of several triangles $\Delta_{k_1}, \ldots, \Delta_{k_n}$ (in general, m depends on w_0). If $f(e)$ is homeomorphic in a neighborhood of e_0, then $m = 1$, and the covering of $\mathscr{N}(w_0)$ reduces to a single "sheet." However, if $f(e)$ fails to be homeomorphic in every neighborhood of e_0, then $m > 1$, and the covering of $\mathscr{N}(w_0)$ consists of m distinct "sheets." In this way, we obtain a kind of "quilt" or "many-sheeted surface," sewn together from triangular patches, which covers part or all of the sphere Σ. This geometric object S serves as a *topological model* of E and at the same time explains further why E is called a *covering* surface.

Specification of a "point" of S requires knowledge not only of the point $w = f(e) \in \Sigma$ but also of the particular inverse image $e \in E$ associated with w (or equivalently, of the unique triangle $T_k \in \mathscr{T}$ to which e belongs). In other words, the points of S are really ordered pairs (e, w), where $e \in E$, $w \in \Sigma$. Every point $s = (e, w) \in S$ with the same value of w is said to "lie over" w, and the mapping $w = p(s)$ carrying the point $s = (e, w) \in S$ into the point $w \in \Sigma$ is called the *projection* of S onto Σ. We can also regard S as a Riemann covering surface, homeomorphic to the original surface E (see Problems 7.16 and 7.17).

If there is a star of triangles $T_1, \ldots, T_n \in \mathscr{T}$ at e_0, and if a neighborhood $w_0 = f(e_0)$ is covered m times by the triangles $\Delta_1 = f(T_1), \ldots, \Delta_n = f(T_n)$, then the point $(e_0, w_0) \in S$ is called a *branch point of order* $m - 1$. Clearly,

---

[40] In writing $\mathscr{T} = \{T_k\}$, we rely upon Theorem 7.9.

S contains no more than countably many branch points, which can have no limit points on S (however, see Problem 7.19).

Things are particularly simple if E is closed:

THEOREM 7.18. *Let E be a closed Riemann covering surface, with topological model S. Then there is an integer m such that S is an m-sheeted covering of the whole sphere Σ.*[41]

Proof. Let $\mathscr{T}$ and $w = f(e)$ be the triangulation and interior mapping associated with E. According to Theorem 7.10, $\mathscr{T}$ consists of a finite number of triangles $T_1, \ldots, T_n$. Let M be the set of possible values of $f(e)$, $e \in E$. Since $f(e)$ is an interior mapping, M is open. Moreover, if $w' \in \Sigma$ is a limit point of M, then, using the compactness of E, we can find two sequences $\{w_k\}$ and $\{e_k\}$, where $w_k = f(e_k)$, such that $\{w_k\}$ approaches w' while $\{e_k\}$ approaches a point $e' \in E$. But then $f(e') = w'$, by continuity of $f(e)$, and hence $w' \in M$. It follows from Theorem I.4.16 that $M = \Sigma$, and hence $f(e)$ maps E *onto* Σ.

Next let $\mathscr{G}$ be the domain obtained from Σ by excluding all points lying under branch points of S (there are only a finite number of such points). Then every point $w \in \mathscr{G}$ has a finite number of inverse images $e \in E$ (no more than one in each triangle $T_1, \ldots, T_n$), and $w = f(e)$ is homeomorphic in a neighborhood of each inverse image. But then, by the same argument as in the proof of Theorem 7.13, where now $T_1 \cup \cdots \cup T_n$ is the whole surface E and not just the union of a star of triangles at e_0, we find that *every* point $w \in \mathscr{G}$ has precisely m inverse images, where m is some integer $\leqslant n$.

36. Regular Analytic Curves

A continuous curve L is said to be *analytic* if its equation can be written in the form

$$z = \lambda(t) \qquad (\alpha \leqslant t \leqslant \beta),$$

where $\lambda(t)$ is an analytic function of the parameter t at every point of the interval $[\alpha, \beta]$. Thus, if L is analytic, $\lambda(t)$ has a power series expansion of the form

$$z = \lambda(t) = \sum_{k=0}^{\infty} (a_k + ib_k)(t - t_0)^k \qquad (a_k, b_k \text{ real}) \tag{7.16}$$

in a neighborhood of every $t_0 \in [\alpha, \beta]$. Taking real and imaginary parts of (7.16), we obtain

$$x = \sum_{k=0}^{\infty} a_k(t - t_0)^k, \qquad y = \sum_{k=0}^{\infty} b_k(t - t_0)^k,$$

[41] I.e., every point $w \in \Sigma$ is covered by precisely m sheets of S (unless w lies under one of the finitely many branch points of S).

i.e., x and y have power series expansions in a neighborhood of every $t_0 \in [\alpha, \beta]$ and hence are themselves analytic in the interval $[\alpha, \beta]$. The curve L is said to be *regular analytic* if, in addition to being analytic, the function $\lambda(t)$ can be chosen to satisfy the condition $\lambda'(t_0) \neq 0$ at every point of $[\alpha, \beta]$. If this condition is satisfied, then every point $t_0 \in [\alpha, \beta]$ has a neighborhood in which the equation of L can be written in the form $y = y(x)$ or $x = x(y)$, depending on whether $x'(t_0) \neq 0$ or $y'(t_0) \neq 0$, where in the first case $y(x)$ is analytic at $x = x(t_0)$ while in the second case $x(y)$ is analytic at $y = y(t_0)$.

Example 1. The straight line segment $z = z_0 + le^{i\theta}t$, the circle $z = z_0 + re^{2\pi it}$ and the ellipse $z = z_0 + a \cos 2\pi t + ib \sin 2\pi t$ (in each case $0 \leqslant t \leqslant 1$) are all regular analytic curves.

Example 2. The curve L with equation

$$z = \lambda(t) = t^3 + it^4 \qquad (-1 \leqslant t \leqslant 1) \tag{7.17}$$

is analytic, but not regular analytic. To see this, let $t = \mu(\tau)$ be any continuous strictly monotonic change of parameter (cf. p. I.60) such that $z = \mu^3(\tau) + i\mu^4(\tau)$ is still analytic and $\mu(0) = 0$ (the latter assumption entails no loss of generality). Then the functions $\mu^3(\tau)$ and $\mu^4(\tau)$ are analytic in a neighborhood of the point $\tau = 0$, and hence so is their ratio [42]

$$\frac{\mu^4(\tau)}{\mu^3(\tau)} = \mu(\tau).$$

It follows that

$$x'(0) = 3\mu^2(\tau)\mu'(\tau)|_{\tau=0} = 0, \qquad y'(0) = 4\mu^3(\tau)\mu'(\tau)|_{\tau=0} = 0,$$

and hence, as asserted, (7.17) cannot be a regular analytic curve. Correspondingly, (7.17) cannot be written in the form $y = y(x)$ where y is analytic at $x = 0$, nor in the form $x = x(y)$ where x is analytic at $y = 0$, since in fact

$$y = x^{4/3}, \qquad x = y^{3/4}.$$

THEOREM 7.19. *Let*

$$L: z = \lambda(t) \qquad (\alpha \leqslant t \leqslant \beta)$$

be a regular analytic curve contained in a domain G, and suppose $w = f(z)$ maps G conformally onto a domain in the w-plane. Then the curve $\Lambda = f(L)$ is regular analytic, i.e., the conformal image of a regular analytic curve is itself regular analytic.

Proof. The curve Λ has equation

$$w = f[\lambda(t)] \qquad (\alpha \leqslant t \leqslant \beta),$$

[42] The *a priori* possibility of $\mu^4(\tau)/\mu^3(\tau)$ having a pole at $\tau = 0$ is precluded by the fact that $\mu(\tau)$ is bounded.

where $f[\lambda(t)]$ is analytic in $[\alpha, \beta]$, being an analytic function of an analytic function. Moreover

$$\frac{dw}{dt} = \frac{df}{dz}\frac{dz}{dt} \neq 0$$

at every point of $[\alpha, \beta]$, since $f(z)$ is univalent on $G \supset L$ and L is regular analytic.

Two circular arcs either intersect in a finite number of points (in fact, 0, 1 or 2 points), or else they have one or two arcs in common. General regular analytic arcs have essentially the same property, as shown by

THEOREM 7.20. *Given two regular analytic arcs*

$$L_1: z = \lambda_1(t') \qquad (\alpha' \leqslant t' \leqslant \beta')$$

and

$$L_2: z = \lambda_2(t'') \qquad (\alpha'' \leqslant t'' \leqslant \beta''),$$

let $M = L_1 \cap L_2$. *Then* M *is either finite (possibly empty), or else* M *consists of one or two arcs whose end points are also end points of* L_1 *and* L_2.[43]

Proof. Let

$$M' = \lambda_1^{-1}(M), \qquad M'' = \lambda_2^{-1}(M),$$

so that M' is the set of values of the parameter t' corresponding to points of intersection of L_1 and L_2, and similarly for M''. If both M' and M'' are finite, then so is M and the theorem is proved. Thus, suppose M' (say) is infinite. Let t_0' be a limit point of M', and let $\{t_n'\}$ be a sequence of distinct points of M' converging to t_0'. Since $\lambda_1'(t_0') \neq 0$ by hypothesis, the analytic function $z = \lambda_1(t')$ maps some (circular) neighborhood $\mathcal{N}(t_0')$ of the point t_0' onto a domain G containing the point $z_0 = \lambda_1(t_0')$. This mapping is one-to-one, and hence distinct values of t_n' contained in $\mathcal{N}(t_0')$ correspond to distinct points $z_n' = \lambda_1(t_n')$ of some arc

$$l_1: z = \lambda_1(t) \qquad (t_0' - \delta \leqslant t' \leqslant t_0' + \delta)$$

contained in both L_1 and G. For each z_n', let t_n'' be a value of the parameter t'' such that $z_n' = \lambda_2(t_n'')$. Then, going over to a subsequence if necessary, we can assume that $\{t_n''\}$ consists of distinct points converging to a point t_0'' (in particular, M'' is infinite). Clearly

$$\lambda_2(t_0'') = \lim_{n\to\infty} \lambda_2(t_n'') = \lim_{n\to\infty} \lambda_1(t_n') = \lambda_1(t_0') = z_0,$$

so that z_0 is a limit point of both curves L_1 and L_2.

[43] If there are two arcs, one may reduce to a single point (see below).

Now let ε be so small that the arc

$$l_2 : z = \lambda_2(t) \qquad (t_0'' - \varepsilon \leqslant t'' \leqslant t_0'' + \varepsilon)$$

($l_2 \subset L_2$) is contained in G. Then l_2 shares infinitely many distinct points $z_n' = \lambda_2(t_n'')$ with l_1. Under the conformal mapping $t' = \lambda_1^{-1}(z)$ of the domain G onto $\mathcal{N}(t_0')$, the arc l_1 goes into the interval $[t_0' - \delta, t_0' + \delta]$, while l_2 goes into a regular analytic curve (recall Theorem 7.19) intersecting this interval of the real axis in infinitely many distinct points t_n' with limit point t_0'. The ordinate $y(t')$ of this curve is an analytic function of t', vanishing at infinitely many distinct points t_n' with limit point t_0'. Therefore $y(t')$ vanishes identically in some closed interval $I \subset \mathcal{N}(t_0')$ with t_0' as an interior point (why?), and hence $\lambda_1^{-1}(l_2)$ coincides with $\lambda_1^{-1}(l_1)$ in I. It follows that L_1 and L_2 share an arc l in some neighborhood of the point $z_0 = \lambda_1(t_0') = \lambda_2(t_0'')$.

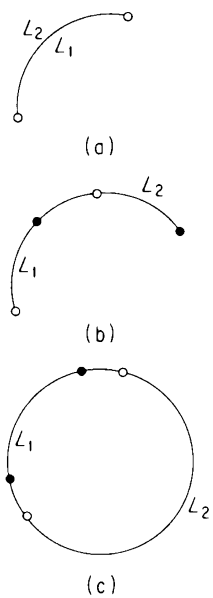


FIGURE 7.7

Applying the same argument to the end points of l (which are both limit points of M'), we continue l as far as possible in both directions, by "splicing on" more common arcs of L_1 and L_2. Clearly, this process of continuation can terminate only when both end points of l finally become end points of L_1 and L_2, as they must in countably many steps (justify both assertions). In fact there are three possible ways in which this can happen, as illustrated schematically in Figure 7.7:

a) Both end points of L_1 coincide with those of L_2. Then L_1 and L_2 coincide as point sets, although they may have opposite orientations.

b) One and only one end point of each curve L_1, L_2 is an interior point of the other. Then L_1 and L_2 do not coincide and they have one arc in common (they may also share an end point).

c) Both end points of each curve L_1, L_2 are interior points of the other curve.[44] Then, in general, L_1 and L_2 do not coincide, and they have two arcs in common. However, if L_1 and L_2 are both closed, they coincide, while if only one of the curves is closed, the two arcs coalesce to form a single arc.

In any case, M consists of one or two arcs (recall footnote 43) whose end points are also end points of L_1 and L_2, and the proof is complete.

---

[44] In this case, after continuing l as far as possible in both directions, it may be found that the set $\tilde{M} = M' - \lambda_1^{-1}(l)$ is still infinite, so that the above construction applied to $\tilde{M}$ leads to another common arc of L_1 and L_2.

37. The Riemann Surface of a Meromorphic Function

LEMMA 1. *Let the function $f(z)$ be analytic at the finite point $z = z_0$, where it has a w_0-point of order $k > 1$. Then there is a star of topological triangles $T_1, \ldots, T_{k+1}$ at z_0, all bounded by regular analytic curves, such that $f(z)$ is homeomorphic on every triangle T_j $(j = 1, \ldots, k + 1)$ and also in a neighborhood of every boundary point of T_j except the point z_0 itself.*

Proof. Let $\mathcal{N}(z_0)$ and $\mathcal{N}(w_0) = \mathcal{N}(w_0; r)$ be the same as in Theorem II.3.4, Corollary 1, so that every point $w \in \mathcal{N}(w_0)$ except w_0 itself has precisely k distinct inverse images in $\mathcal{N}(z_0)$. The Taylor series expansion of $f(z)$ at z_0 is of the form

$$w = f(z) = w_0 + a_k(z - z_0)^k + a_{k+1}(z - z_0)^{k+1} + \cdots \qquad (a_k \neq 0). \tag{7.18}$$

As in the proof of Theorem II.3.6, we write

$$w = f(z) = w_0 + t^k \tag{7.19}$$

and

$$t = \varphi(z) = \sqrt[k]{a_k}(z - z_0)\left[1 + \frac{a_{k+1}}{a_k}(z - z_0) + \cdots\right]_s^{1/k}, \tag{7.20}$$

where $\sqrt[k]{a_k}$ denotes any fixed kth root of a_k, and the subscript s denotes any fixed single-valued branch of the expression to which it is attached. The function (7.20) has a simple zero at z_0 and no other zeros in $\mathcal{N}(z_0)$, and hence every point in the disk $|t| < \sqrt[k]{r}$ has a unique inverse image in $\mathcal{N}(z_0)$. Let K be the disk $|t| < \rho$, where $\rho > 0$ is any number less than $\sqrt[k]{r}$. If $D = \varphi^{-1}(K)$, then D is a Jordan domain contained in $\mathcal{N}(z_0)$, and the boundary of D is a regular analytic curve (by Theorem 7.19).

We now divide $\bar{K}$ into $k + 1$ closed sectors (topological triangles) by drawing $k + 1$ radii from the point $t = 0$, with inclinations

$$\frac{2j\pi}{k + 1} \qquad (j = 0, 1, \ldots, k).$$

The inverse images of these sectors under the mapping $t = \varphi^{-1}(z)$ form a set of topological triangles $T_1, \ldots, T_{k+1}$ with common vertex z_0, such that

$$T_1 \cup \cdots \cup T_{k+1} = \bar{D}.$$

In fact, $T_1, \ldots, T_{k+1}$ form a star of triangles at z_0, in the sense of Theorem 7.7,[45] where the sides of each triangle are regular analytic curves. The

[45] At this stage, we are not concerned with the rest of the triangulation.

rest of the lemma is an immediate consequence of the construction of the triangles $T_1, \ldots, T_{k+1}$, the fact that $\varphi(z)$ is univalent on $\bar{D}$ and the nature of the mapping (7.19).

Remark. The triangles in the star $T_1, \ldots, T_{k+1}$ can be made to lie in any neighborhood $\mathcal{N}(z_0)$, however small. In fact, it is only necessary to choose the number ρ sufficiently small.

LEMMA 2. *Lemma* 1 *remains valid if* $z_0 = \infty$.

Proof. Recall the proof of Theorem II.3.4, Corollary 4.

LEMMA 3. *Lemma* 1 *remains valid if* $w_0 = \infty$, *i.e., if* $f(z)$ *has a pole of order* $k > 1$ *at* z_0.

Proof. In this case, we have

$$w = f(z) = a_{-k}(z - z_0)^{-k} + a_{-k+1}(z - z_0)^{-k+1} + \cdots \qquad (a_{-k} \neq 0)$$

instead of (7.18), and the situation is handled as in the proof of Theorem II.3.4. Corollary 5.

The key role played by interior mappings and Riemann covering surfaces in the theory of functions is revealed by the following theorem, where the underlying topological surface E is a domain in the complex plane:

THEOREM 7.21. *Let* G *be a domain in the extended* z*-plane, and let* $w = f(z) \not\equiv$ const *be meromorphic on* G. *Then* $w = f(z)$ *is an interior mapping of* G *into the sphere (extended* w*-plane). In particular,* G *becomes a Riemann covering surface when equipped with the mapping* $w = f(z)$.

Proof. The function $f(z)$ has only countably many multiple A-points $z_1, \ldots, z_n, \ldots$ (including poles), of orders $k_1, \ldots, k_n, \ldots$ At each point z_n, we construct a closed domain $\bar{D}_n$, equal to the union of a star of $k_n + 1$ triangles at z_n with the properties prescribed by Lemmas 1–3, making sure that $\bar{D}_m \cap \bar{D}_n = 0$ if $m \neq n$ (it follows from the remark after Lemma 1 that this is always possible). Every other point $z \in G$ is a simple A-point, and hence has a closed neighborhood $\overline{\mathcal{N}(z)}$ in which $f(z)$ is univalent (cf. Theorem II.4.1). In this way, with every point $z \in G$ we associate a closed domain containing z as an interior point.

Now let $\{F_n\}$ be an increasing sequence of compact sets approximating G, in the sense of p. I.373. By the Heine-Borel theorem, F_1 can be covered by a finite number of domains of the type $\overline{\mathcal{N}(z)}$ or $\bar{D}_n$, and similarly for $F_2 - F_1$, $F_3 - F_2, \ldots$ This leads to a countable family $\mathcal{M}$ of closed domains covering G itself, such that every compact subset of G is covered by a finite subfamily of $\mathcal{M}$. Without loss of generality, we can assume that no domain in $\mathcal{M}$ is a proper subset of another domain in $\mathcal{M}$. Let $\mathcal{M}'$ be the countable family of closed domains (disks and triangles) obtained by replacing every domain $\bar{D}_n \in \mathcal{M}$ by the $k_n + 1$ triangles of which $\bar{D}_n$ is the union. Then $f(z)$ is homeomorphic

on every domain $M \in \mathscr{M}'$ and also in a neighborhood of every boundary point of M except possibly vertices if M is a triangle. The boundary of every domain in $\mathscr{M}'$ consists of a finite number of regular analytic curves (one for a disk and three for a triangle). It follows from Theorem 7.20 that a given domain in $\mathscr{M}'$ can intersect only a finite number of other domains in $\mathscr{M}'$. Therefore the domains in $\mathscr{M}'$ are partitioned by their boundaries (which intersect, in general) into a countable family $\mathscr{M}''$ of closed Jordan domains, *with disjoint interiors*, consisting (so to speak) of the pieces into which each domain in $\mathscr{M}'$ is cut by its intersections with the other domains in $\mathscr{M}'$ (no piece common to several domains in $\mathscr{M}'$ is counted more than once).[46] As before, G is covered by $\mathscr{M}''$, and every compact subset of G is covered by a finite subfamily of $\mathscr{M}''$. Moreover, $f(z)$ is homeomorphic on every domain $M \in \mathscr{M}''$ and also in a neighborhood of every boundary point of M except the points $z_1, \ldots, z_n, \ldots$

We must still convert $\mathscr{M}''$ into a triangulation $\mathscr{T}$ of the "surface" G, by further subdividing the domains in $\mathscr{M}''$. This is done as follows: The boundary of every domain in $\mathscr{M}''$ is shared with a finite number of other domains in $\mathscr{M}''$, and hence consists of a finite number of arcs, each a connected component of the common boundary of a pair of "adjacent" domains. The end points of these arcs will be called "vertices" of the family $\mathscr{M}''$. A domain in $\mathscr{M}''$ may have only two vertices, and then we introduce an extra point on one of its sides, which we also call a vertex. Given any domain $M \in \mathscr{M}''$ with vertices $\zeta_1, \ldots, \zeta_m$, we now choose a point $\zeta_0 \in \text{int}\,(M)$ and join ζ_0 to $\zeta_1, \ldots, \zeta_m$ by m Jordan arcs $l_1, \ldots, l_m$ such that $l_j \subset \text{int}\,(M)$ except for the point ζ_j, while l_j and l_k $(j \neq k)$ have only the point ζ_0 in common. This leads to a star of triangles at ζ_0, whose union equals M. Suppose we replace M by this star of triangles, and similarly for every other domain in $\mathscr{M}''$, thereby obtaining a triangulation $\mathscr{T}$ of G. Then $f(z)$ is obviously an interior mapping relative to $\mathscr{T}$, and the proof is complete.

COROLLARY. *Let $G, f(z)$ and $\mathscr{T}$ be the same as in Theorem* 7.21 *and its proof, and let z_0 be a w_0-point of order $k > 1$ of $f(z)$. Then there is a star of triangles $T_1, \ldots, T_\nu \in \mathscr{T}$ at z_0 such that the triangles $\Delta_1 = f(T_1), \ldots, \Delta_\nu = f(T_\nu)$ cover a neighborhood $\mathscr{N}(w_0)$ k times.*

Proof. Recall Theorem 7.13, Corollary 1.

DEFINITION. *The Riemann covering surface figuring in Theorem* 7.21 *(or its topological model) is called the Riemann surface of the function $w = f(z)$, or of the inverse function $z = f^{-1}(w)$.*[47]

---

[46] We can think of G as a "jigsaw puzzle" with the domains in $\mathscr{M}''$ as the "pieces," or as a "mosaic" with the domains in $\mathscr{M}''$ as the "tiles."

[47] The latter designation is particularly apt in certain situations, as in Theorem 7.23 below.

THEOREM 7.22. *Two nonconstant functions $f(z)$ and $g(z)$ meromorphic on the same domain G have equivalent Riemann surfaces if and only if there exists a function $\Phi(z)$ mapping G conformally onto itself such that*

$$f(z) = g[\Phi(z)], \qquad g(z) = f[\Phi^{-1}(z)]. \tag{7.21}$$

Proof. According to Definition 3, p. 238, the two Riemann surfaces are equivalent if and only if there exists a homeomorphism $\Phi(z)$ of G onto itself such that (7.21) holds. But then

$$\Phi(z) = g^{-1}[f(z)], \qquad \Phi^{-1}(z) = f^{-1}[g(z)],$$

which implies that $f(z)$ and $g(z)$ are both univalent on G (why?). It follows that $\Phi(z)$ itself is univalent and hence conformal on G.

Except for the trivial case where $w = f(z)$ is univalent on G, the inverse function $z = f^{-1}(w)$ is multiple-valued on the domain $f(G) \subset \Sigma$, where Σ is the sphere (extended w-plane). However, the inverse function becomes single-valued when defined "in the natural way" on S, the topological model of the Riemann surface of $z = f^{-1}(w)$:

THEOREM 7.23. *Let $z = f^{-1}(w)$ be the inverse of a nonconstant function meromorphic on a domain G, and let S be the topological model of the Riemann surface of $f^{-1}(w)$. Moreover, let $w = p(s)$ be the projection of S onto the sphere Σ, as defined on p. 240. Then the function*

$$z = f^{-1}[p(s)] \tag{7.22}$$

is single-valued on S.

Proof. We must show that distinct points z', $z'' \in G$ have distinct inverse images under the mapping (7.22). But these inverse images are just

$$s' = (z', w'), \qquad s'' = (z'', w''),$$

where $w' = f(z')$, $w'' = f(z'')$, and hence are distinct if $z' \neq z''$, even if $w' = w''$.

Remark 1. It follows from the corollary to Theorem 7.21 that if $f(z)$ has a w_0-point of order $k > 1$ at z_0, then (z_0, w_0) is a branch point of order $k - 1$ of S, as defined on p. 240. Moreover, every branch point of S lies over a multiple A-point of $f(z)$, since every other point of G has a neighborhood in which $f(z)$ is homeomorphic.

Remark 2. Suppose $f(z)$ has a w_0-point of order $k > 1$ at z_0, so that a neighborhood of $\mathcal{N}(w_0)$ is covered k times by the images $\Delta_1 = f(T_1), \ldots, \Delta_\nu = f(T_\nu)$ of a star of triangles $T_1, \ldots, T_\nu \in \mathcal{T}$ at z_0. Then, given any point $w \in \mathcal{N}'(w_0)$, we denote the k distinct inverse images of w by $z^{(1)}, \ldots, z^{(k)}$, arranged in the order of their appearance in the star $T_1, \ldots, T_\nu$. More

exactly, if $z^{(j)} \in T_{m_j}$, $z^{(j')} \in T_{m_{j'}}$, then $j < j'$ implies $m_j < m_{j'}$ and conversely. Thus w lies under k distinct points

$$(z^{(1)}, w), \ldots, (z^{(k)}, w)$$

of S, where $(z^{(j)}, w)$ is said to lie on the jth sheet of S. Let L be any Jordan curve contained in $\mathcal{N}(w_0)$ except for its end points, one at w_0 and the other at a boundary point of $\mathcal{N}(w_0)$. Then, given any domain $\mathcal{G} \subset \mathcal{N}(w_0) - L$, the function $f_j^{-1}(w)$ assigning w the value $z^{(j)}$ on the jth sheet of S defines a single-valued branch of $f^{-1}(w)$ on $\mathcal{G}$ (in the sense of p. I.213), i.e., each sheet of S corresponds to a single-valued branch of $f^{-1}(w)$. Thus the Riemann surface of a multiple-valued function is the natural tool for describing all the single-valued branches of the function and the relation between the various branches.

It is easy to see that the point w_0 is an algebraic branch point of order $k - 1$, as defined on p. I.217 in connection with the elementary theory of multiple-valued functions. In fact, if $C \subset \mathcal{N}(w_0)$ is any closed Jordan curve such that $w_0 \in I(C)$, then one circuit around C in a fixed direction carries every (single-valued) branch of $f^{-1}(w)$ into another branch, while k circuits around C (*and no fewer*) carry every branch of $f^{-1}(w)$ into itself (cf. p. II.94). Thus the concept of a branch point of the Riemann surface S is intimately related to (and serves as a geometric model of) the previously introduced concept of an (algebraic) branch point of the multiple-valued function $z = f^{-1}(w)$.

Remark 3. Since S is homeomorphic to G, all the information about the function $z = f^{-1}(w)$ provided by S can also be deduced from the domain G itself, equipped with a triangulation $\mathcal{T}$, as in the proof of Theorem 7.21. Clearly, the interior of every triangle $T \in \mathcal{T}$ is a domain of univalence of the function $w = f(z)$, in the sense of p. I.213.[48] On the other hand, in our earlier studies of multiple-valued functions (specifically, those obtained by taking the inverses of certain elementary entire and meromorphic functions), we partitioned G into domains of univalence which were not required to form a triangulation. In fact, once it has been established that a given meromorphic function $w = f(z)$ is an interior mapping (relative to a suitable triangulation $\mathcal{T}$), there is no need to study the inverse function $z = f^{-1}(w)$ by using the triangular domains of univalence corresponding to $\mathcal{T}$, and any other partition of G into countably many domains of univalence is equally suitable, even for the purpose of constructing a model of the Riemann surface of $z = f^{-1}(w)$. Thus, regarding G as a "mosaic" (cf. footnote 46, p. 247), we

[48] To complete the picture, it should be noted at this point that the function $f(z)$ figuring in Remark 2, p. I.213, described as "wide-sense continuous on a domain G, and analytic on G except possibly on a set $I \subset G$ consisting entirely of isolated points at which $f(z)$ approaches ∞," is precisely what is meant by a function meromorphic on G (see Sec. 10).

can assemble G from "tiles" of arbitrary shape, provided only that $f(z)$ is univalent on (the interior of) every tile. Then, continuing to denote the tiles by $T_1, \ldots, T_n, \ldots$ and their images by $\Delta_1 = f(T_1), \ldots, \Delta_n = f(T_n), \ldots$, we can construct a model of the Riemann covering surface S from the family of regions $\{\Delta_n\}$ by using the following prescription: Every pair of regions Δ_k, Δ_l $(k \neq l)$ sharing common "pieces of boundary" is pasted together along the common pieces.

38. Examples

We now give several examples illustrating the theory developed in the preceding section. As in Chap. I.11 (Sec. I.54 *et seq.*), we use notation such that $w = f(z)$, rather than $z = f^{-1}(w)$, is the multiple-valued function.

Example 1.[49] Let $n > 1$ be an integer, and consider the function

$$w = f(z) = \sqrt[n]{z}, \tag{7.23}$$

which is the inverse of the function $z = w^n$. Let $\mathscr{G}$ be the extended w-plane, and draw $2n$ rays from the point $w = 0$ such that the first ray is the positive real axis and the angles between adjacent rays all equal π/n. Let $\mathscr{G}_j$ $(j = 1, \ldots, 2n)$ be the interior of the angle bounded by the rays with inclinations $(j-1)\pi/n$ and $j\pi/n$. Then the function $z = w^n$ is univalent on every closed domain $\overline{\mathscr{G}}_j$. The image $\Delta_j = f^{-1}(\overline{\mathscr{G}}_j)$ is either the closed upper half-plane or the closed lower half-plane, depending on whether j is odd or even. Hence, to construct a model of the Riemann surface of the function (7.23), we need only paste these half-planes together along their boundaries in the same way as the domains $\overline{\mathscr{G}}_j$ themselves are pasted together. Thus, denoting the part of the boundary of Δ_j consisting of the positive real axis by j_+ and the part consisting of the negative real axis by j_-, we paste 1_- to 2_-, 2_+ to 3_+, 3_- to $4_-, \ldots, (2n-1)_-$ to $(2n)_-$, and finally $(2n)_+$ to 1_+. The corresponding system of "pasted sheets" is shown in Figure 7.8, for the case $n = 4$. As a result, we obtain an n-sheeted covering S of the whole sphere, representing the Riemann surfaces of the function $w = \sqrt[n]{z}$. The surface S has two branch points of order $n-1$, lying over the points $z = 0$ and $z = \infty$. The function $w = \sqrt[n]{z}$ is single-valued and analytic at every point of S

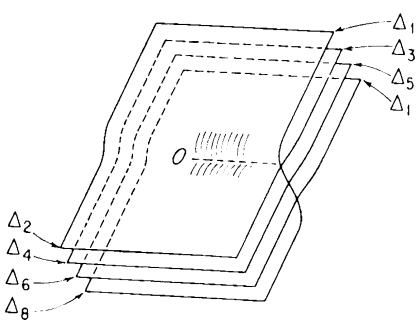


FIGURE 7.8

[49] Cf. Sec. I.54.

except at the branch point lying over $z = \infty$ where it has a pole (recall Theorem 7.16).

Example 2.[50] Consider the function

$$w = f(z) = \operatorname{Ln} z, \tag{7.24}$$

which is the inverse of the function $z = e^w$. Let $\mathcal{G}$ be the finite w-plane, which we partition into infinitely many open strips

$$\mathcal{G}_j: j\pi < u < (j+1)\pi \qquad (j = 0, \pm 1, \pm 2, \ldots),$$

where $u = \operatorname{Re} w$. Then the function $z = e^w$ is univalent on every strip $T_j = \overline{\mathcal{G}}_j - \{\infty\}$.[51] The image $\Delta_j = f^{-1}(T_j)$ is either the closed upper half-plane minus the points 0 and ∞ or the closed lower half-plane minus the points 0 and ∞, depending on whether j is even or odd. Hence, to construct a model of the Riemann surface of the function (7.24), we need only paste this countably infinite collection of half-planes together along their boundaries in the same way as the strips T_j themselves are pasted together. The corresponding system of pasted sheets is shown in Figure 7.9, and is analogous to the system in Figure 7.8, except that there are now infinitely many regions Δ_j, with no first or last region, and correspondingly no final pasting together of a first and a last sheet of the covering surface. As a result, we obtain an "infinite-sheeted" covering S of the whole sphere, except for the points $z = 0$ and $z = \infty$ which lie under no points of S whatsoever. The surface S has no branch points at all, since the function $z = e^w$ is univalent in a neighborhood of every point of $\mathcal{G}$. The function $w = \operatorname{Ln} z$ is single-valued and analytic at every point of S without exception.

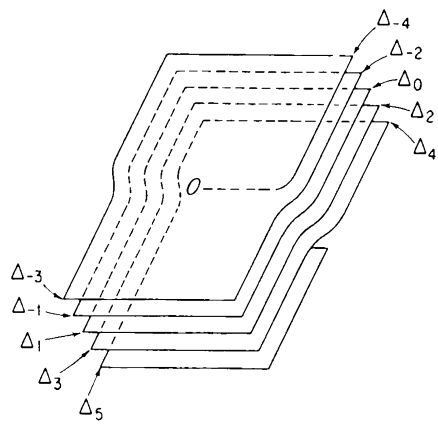


FIGURE 7.9

Example 3. Consider the function

$$w = f(z) = \operatorname{Arc\,sin} z, \tag{7.25}$$

which is the inverse of the function $z = \sin w$. Let $\mathcal{G}$ be the finite w-plane, which we partition into infinitely many open half-strips

$$\mathcal{G}'_j: (j - \tfrac{1}{2})\pi < u < (j + \tfrac{1}{2})\pi, \quad v > 0 \qquad (j = 0, \pm 1, \pm 2, \ldots)$$

---

[50] Cf. Sec. I.56.

[51] We exclude the point $w = \infty$ at which e^w has no limit, finite or infinite (then $T_j \subset \mathcal{G}$).

and

$$\mathcal{G}_j'':(j - \tfrac{1}{2})\pi < u < (j + \tfrac{1}{2})\pi, \quad v < 0 \qquad (j = 0, \pm 1, \pm 2, \ldots),$$

where $w = u + iv$. Then the function $z = \sin w$ is univalent on every half-strip $T_j' = \mathcal{G}_j' - \{\infty\}$, $T_j'' = \mathcal{G}_j'' - \{\infty\}$. Let Π_U and Π_L denote the open upper and lower half-planes $\operatorname{Im} w > 0$ and $\operatorname{Im} w < 0$, respectively. Then the image $\Delta_j' = f^{-1}(T_j')$ is either $\overline{\Pi}_U - \{\infty\}$ or $\overline{\Pi}_L - \{\infty\}$, depending on whether j is even or odd, while $\Delta_j'' = f^{-1}(T_j'')$ is either $\overline{\Pi}_U - \{\infty\}$ or $\overline{\Pi}_L - \{\infty\}$, depending on whether j is odd or even. Hence, to construct a model of the Riemann surface of the function (7.25), we need only paste this countably infinite collection of half-planes together along their boundaries in the same way as the half-stripes T_j' and T_j'' themselves are pasted together. As a result,

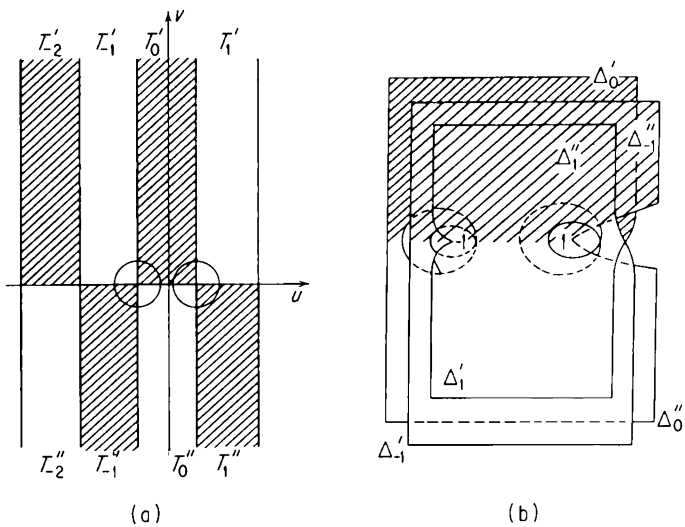


FIGURE 7.10

we obtain an infinite-sheeted covering S of the whole sphere, except for the point $z = \infty$ which lies under no point of S whatsoever. Since $\sin w$ never becomes infinite on $\mathcal{G}$, while its derivative $\cos w$ vanishes only when

$$w = \frac{\pi}{2} + n\pi \qquad (n = 0, \pm 1, \pm 2, \ldots),$$

all the branch points of S lie over the points

$$z = \sin\left(\frac{\pi}{2} + n\pi\right) = \pm 1$$

of the sphere. There are infinitely many such branch points, each of order 1 (why?). Figure 7.10 shows the two branch points which are the images of the points $w = -\pi/2$ and $w = \pi/2$. The function $w = \operatorname{Arc} \sin z$ is single-valued and analytic at every point of S without exception.

Example 4. Finally consider the elliptic integral

$$w = f(z) = \int_{\infty}^{z} \frac{dt}{\sqrt{4t^3 - g_2 t - g_3}} \tag{7.26}$$

which is the inverse of Weierstrass' elliptic function $z = \wp(w)$. For simplicity, let

$$\wp(w) = \wp(w \mid \alpha, i\beta),$$

as in Sec. 22. Let $\mathscr{G}$ be the extended w-plane, which we partition into infinitely many open rectangles

$$\mathscr{G}_{jk}: j\alpha < u < (j+1)\alpha, \quad k\beta < v < (k+1)\beta \qquad (j, k = 0, \pm 1, \pm 2, \ldots),$$

where $w = u + iv$. Then, according to Sec. 22, the function $z = \wp(w)$ is univalent on every closed rectangle $\overline{\mathscr{G}}_{jk}$. Suppose we shade "every other" rectangle $\overline{\mathscr{G}}_{jk}$ in checkerboard fashion, as shown in Figure 7.11(a), with

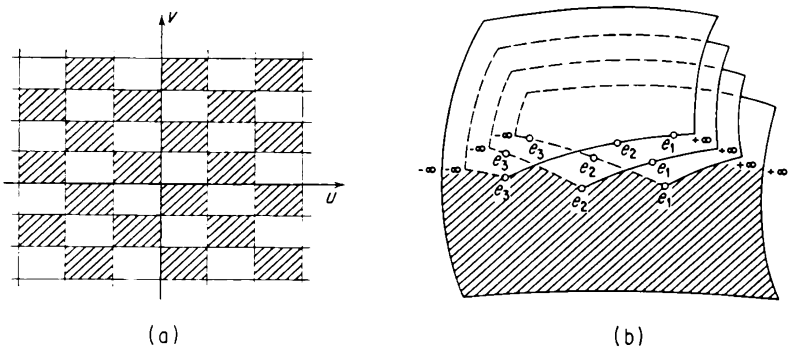


FIGURE 7.11

$\overline{\mathscr{G}}_{00}$ as one of the shaded rectangles. Then the image $\Delta_{jk} = f^{-1}(\overline{\mathscr{G}}_{jk})$ is either the closed upper half-plane or the closed lower half-plane, depending on whether $\overline{\mathscr{G}}_{jk}$ is unshaded or shaded. (In particular, the vertices 0, α, $\alpha + i\beta$ and $i\beta$ of $\overline{\mathscr{G}}_{00}$ are mapped into the points ∞, e_1, e_2 and e_3 of the real axis.) Hence, to construct a model of the Riemann surface of the function (7.26), we need only paste this countably infinite collection of half-planes together along their boundaries in the same way as the rectangles $\overline{\mathscr{G}}_{jk}$ themselves are pasted together. Figure 7.11(b) shows how this is done for the rectangle $\overline{\mathscr{G}}_{00}$ and the four adjacent unshaded rectangles. As a result, we obtain an infinite-sheeted covering S of the whole sphere. The surface S has infinitely many branch points, each of order 1,[52] lying over the points ∞, e_1, e_2 and e_3 of the sphere. The function (7.26) is single-valued and analytic at every point of S without exception.

[52] Recall from Sec. 21 that the poles of $\wp(w)$ are double, while the zeros of $\wp'(w)$ are simple. It follows from the periodicity of $\wp(w)$ that each of the values ∞, e_1, e_2, e_3 is taken at a countable infinity of distinct points.

PROBLEMS

7.1. Give appropriate definitions of the following topological concepts for an arbitrary Hausdorff space:

a) Closed set; b) Domain; c) Closure of a set; d) Boundary of a set; e) Continuous curve; f) Continuum.

7.2. By a *metric space* we mean a set E together with a nonnegative function $\rho(e, e')$ defined for all $e, e' \in E$ such that

1. $\rho(e, e') = 0$ if and only if $e = e'$;
2. $\rho(e, e') = \rho(e', e)$ for all $e, e' \in E$;
3. $\rho(e, e'') \leqslant \rho(e, e') + \rho(e', e'')$ for all $e, e', e'' \in E$.

Given any $\varepsilon > 0$ and any $e \in E$, the set of all points $e' \in E$ such that $\rho(e, e') < \varepsilon$ is called a *neighborhood* of e. Prove that E is a Hausdorff space (with these neighborhoods).

7.3. Let E be a Hausdorff space, with neighborhoods $\{\mathcal{N}(e)\}$. Prove that any set $F \subset E$ is itself a Hausdorff space when equipped with neighborhoods $\{F \cap \mathcal{N}(e)\}$, $e \in F$.

Comment. F is said to have the *relative topology* induced by E.

7.4. Is the cone $\xi^2 + \eta^2 = \zeta^2$ (with the relative topology induced by three-dimensional Euclidean space) a topological surface? Is it a manifold? How about the half-cone $\xi^2 + \eta^2 = \zeta^2$, $\zeta \geqslant 0$?

7.5. With the same notation as in the remark on p. 218, prove that the mapping $z = \varphi_{e_0}(e)$ carries every closed subset $M \subset \mathcal{N}(e_0)$ into a closed subset of G_{e_0}.

7.6. What is the analogue of the representation (7.1) for a Möbius strip?

7.7. Explain the sense in which Definition 2, p. 223 is invariant under changes of local parameters.

7.8. Generalize Theorem 7.4 to the case of domains E and E^* in the extended complex plane.

7.9. Prove that a triangulable two-dimensional manifold must have a countable base.

7.10. Triangulate

a) The Euclidean plane;
b) The open square $0 < x < 1$, $0 < y < 1$;
c) An arbitrary simply connected domain whose boundary contains more than one point.

7.11. Given a topological space E with a triangulation $\mathcal{T}$, prove that E is orientable (in the sense of Definition 5, p. 231) if and only if it is possible to orient all the triangles in $\mathcal{T}$ in such a way that every pair of adjacent triangles is coherently oriented.

7.12. Assign orientations (in the sense of the preceding problem) to suitable triangulations of the plane, sphere and torus, thereby proving that these surfaces are orientable.

7.13. Prove that the projective plane (introduced on p. I.84) is nonorientable.

7.14. Suppose int (T) is replaced by T in the first condition figuring in the definition of an interior mapping on p. 232. Show that this excludes the case $m = n$ in Theorem 7.13 and its corollaries.

7.15. Let E and E^* be the same as in the example on p. 239, but this time let the interior mappings be

$$w = f(e) = \left(\frac{e}{e-1}\right)^2, \qquad w = f^*(e^*) = e^{*2}.$$

Prove that E and E^* are equivalent Riemann covering surfaces.

Hint. Consider the homeomorphism

$$e^* = \Phi(e) = \frac{e}{e-1}.$$

7.16. Let E, S and $w = f(e)$ be the same as on p. 240. Suppose that by a neighborhood of any point $s_0 = (e_0, w_0) \in S$ we mean the set of all points $s = (e, w)$ such that $e \in \mathcal{N}(e_0)$, where $\mathcal{N}(e_0)$ is any neighborhood of the point $e_0 \in E$. Prove that S is then a topological surface homeomorphic to E.

7.17. Let S be the same topological surface as in the preceding problem. Prove that the projection $w = p(s)$ introduced on p. 240 is an interior mapping of S into the sphere Σ.

7.18. What is the appropriate way to measure angles between curves on the surface S of Problems 7.16 and 7.17?

Hint. Project the curves onto the sphere Σ. The answer depends on whether or not the vertex of the angle is a branch point of S.

7.19. Let S be the same as in the preceding problem. Construct an example where the projections of the branch points on S are dense in Σ.[53]

7.20. Is the curve with equation

$$z = \lambda(t) = t^2 + it^4 \qquad (-1 \leqslant t \leqslant 1)$$

a regular analytic curve?

7.21. Prove that Theorem 7.21 is an immediate consequence of the result cited in Remark 2, p. 232.

7.22. Make a detailed study (as in the examples given in Sec. 38) of the Riemann surfaces of the following functions or of their inverses (n is a positive integer):

a) $z = \left(\frac{w-a}{w-b}\right)^n$; b) $z = \frac{1}{2}\left(w + \frac{1}{w}\right)$; c) $z = \frac{w}{(1-w)^2}$;

d) $z = \frac{1}{2}\left(w^n + \frac{1}{w^n}\right)$; e) $z = \frac{w}{(1+w^n)^2}$.

In particular, find domains of univalence for each function.

---

[53] A set $A \subset B$ is said to be *dense in* B if $\bar{A} \supset B$.

7.23. Study the Riemann surfaces of the following multiple-valued functions:
a) $w = \operatorname{Arc}\cos z$; b) $w = \operatorname{Arc}\tan z$; c) $w = \operatorname{Arc}\sinh z$.

7.24. Study the Riemann surface of the function

$$z = \left(1 + \frac{w}{n}\right)^n$$

as a function of n, showing how it turns into the Riemann surface of $w = \operatorname{Ln} z$ as $n \to \infty$.

Hint. Recall the remark at the end of Sec. I.39.

8

ANALYTIC CONTINUATION

39. Elements. The Complete Analytic Function

We begin by generalizing the concept of a circular element, introduced in Sec. I.84:

DEFINITION 1. *A set $\{G, f(z)\}$ consisting of a domain G in the extended z-plane and a single-valued analytic function $f(z)$ defined on G is called a (function) element, and G is called the domain of the element. Two elements $\{G, f(z)\}$ and $\{D, \varphi(z)\}$ are equal if and only if $G = D, f(z) \equiv \varphi(z)$.*

DEFINITION 2. *Each of two elements $\{G, f(z)\}$ and $\{D, \varphi(z)\}$ is said to be a direct analytic continuation of the other if $G \cap D \neq 0$ and if there exists a domain $g \subset G \cap D$ such that $f(z) = \varphi(z)$ for all $z \in g$.*

Remark. Let $\{G, f(z)\}$, $\{D, \varphi(z)\}$ and g be the same as in Definition 2, and let g' be the connected component of $G \cap D$ containing g. Then, by the interior uniqueness theorem (Theorem I.17.1), $f(z)$ and $\varphi(z)$ coincide on g', and hence on $G \cap D$ itself if $G \cap D$ is connected. However, $f(z)$ and $Q(z)$ may not coincide on the other components of $G \cap D$ if $G \cap D$ is disconnected.

Example 1. Let G be the domain bounded by the ray $-\infty < x \leqslant 0$, $y = 0$ and D the domain bounded by the ray $0 \leqslant x < +\infty$, $y = 0$. Let

$$f(z) = \ln|z| + i\theta, \qquad \varphi(z) = \ln|z| + i\alpha, \tag{8.1}$$

where θ is the value of $\operatorname{Arg} z$ satisfying the condition $-\pi < \theta < \pi$ and α is the value satisfying the condition $0 < \alpha < 2\pi$. Each of the functions (8.1) is single-valued and analytic on its domain of definition. The domain $G \cap D$ is nonempty and consists of two components, the upper half-plane Π_U and

the lower half-plane Π_L. Moreover $\alpha = \theta$ in Π_U while $\alpha = \theta + 2\pi$ in Π_L, and hence $f(z) = \varphi(z)$ if $z \in \Pi_U$ but $f(z) \neq \varphi(z)$ if $z \in \Pi_L$. Therefore $\{G, f(z)\}$ and $\{D, \varphi(z)\}$ are direct analytic continuations of each other (with Π_U as the domain g figuring in Definition 2).

Example 2. Let G, D, $f(z)$ and $\varphi(z)$ be the same as in Example 1, and consider another function

$$\psi(z) = \ln |z| + i\beta$$

defined on D, where β is the value of Arg z satisfying the condition $-2\pi < \beta < 0$. Then $\alpha = \beta + 2\pi$ at every point of D, and hence $\varphi(z) \neq \psi(z)$ if $z \in D$. Therefore $\{D, \varphi(z)\}$ and $\{D, \psi(z)\}$ are not direct analytic continuations of each other. On the other hand, $\beta = \theta$ in the lower half-plane Π_L, and hence $f(z) = \psi(z)$ if $z \in \Pi_L$, so that $\{G, f(z)\}$ and $\{D, \psi(z)\}$ are direct analytic continuations of each other.

DEFINITION 3. *Let*

$$\{G_0, f_0(z)\}, \{G_1, f_1(z)\}, \ldots, \{G_n, f_n(z)\} \tag{8.2}$$

be a set of elements such that $\{G_k, f_k(z)\}$ is a direct analytic continuation of $\{G_{k-1}, f_{k-1}(z)\}$ for every $k = 1, \ldots, n$. Then (8.2) *is called a chain of elements joining $\{G_0, f_0(z)\}$ and $\{G_n, f_n(z)\}$. Each of the elements $\{G_0, f_0(z)\}$ and $\{G_n, f_n(z)\}$ is called an analytic continuation of the other.*[1]

Definition 3 immediately implies

THEOREM 8.1. *The relation leading from an element $\{G, f(z)\}$ to any of its analytic continuations is reflexive, symmetric and transitive.*[2] *More exactly,*

1. *$\{G, f\}$ is an analytic continuation of itself;*[3]
2. *If $\{D, \varphi\}$ is an analytic continuation of $\{G, f\}$, then $\{G, f\}$ is an analytic continuation of $\{D, \varphi\}$;*
3. *If $\{D, \varphi\}$ is an analytic continuation of $\{G, f\}$ and if $\{E, \psi\}$ is an analytic continuation of $\{D, \varphi\}$, then $\{E, \psi\}$ is an analytic continuation of $\{G, f\}$.*

Example. Let G_k be the domain

$$k\pi < \text{Arg } z < (k + 2)\pi \qquad (k = 0, \pm 1, \pm 2, \ldots), \tag{8.3}$$

and let $f_k(z)$ be the function

$$f_k(z) = \ln |z| + i\theta_k,$$

---

[1] The absence of the adjective "direct" is crucial here.

[2] See e.g., G. Birkhoff and S. MacLane, *op. cit.*, p. 30.

[3] $\{G, f\}$ is shorthand for $\{G, f(z)\}$, and similarly for $\{D, \varphi\}$ and $\{E, \psi\}$.

where $\theta_k = \theta_k(z)$ is the value of Arg z satisfying (8.3). Then any two elements $\{G_m, f_m\}$ and $\{G_n, f_n\}$ are analytic continuations of each other. In fact, suppose $n = m + p$, where $p > 0$ (say). Then $\{G_m, f_m\}, \ldots, \{G_n, f_n\}$ is a chain of $p + 1$ elements joining $\{G_m, f_m\}$ and $\{G_n, f_n\}$.

Now let $\mathscr{E}$ be the set of *all* elements, i.e., the set of all possible single-valued analytic functions defined on all possible domains in the extended z-plane. Then $\mathscr{E}$ can be divided into disjoint classes such that two elements belong to the same class if and only if they are analytic continuations of each other. Obviously, every element belongs to some class, and any element $\{G, f\}$ in a given class $\mathscr{K} \subset \mathscr{E}$ uniquely characterizes $\mathscr{K}$ in the sense that $\{D, \varphi\} \in \mathscr{K}$ if and only if $\{D, \varphi\}$ and $\{G, f\}$ are analytic continuations of each other.

DEFINITION 4. *Every class $\mathscr{K}$ of the type just described is said to define a complete analytic function $F(z)$, and distinct classes define distinct complete analytic functions. Every element $\{G, f\}$ is called an element of the function $F(z)$ or a single-valued branch of $F(z)$ [on the domain G].*[4] *The values of the elements of $F(z)$ are called the values of $F(z)$.*

Remark 1. A complete analytic function is in general multiple-valued (see Example 3, pp. 261–263).

Remark 2. Clearly, there is no loss of generality in Definition 4 if it is assumed that the domains G appearing in the elements $\{G, f\} \in \mathscr{K}$ all belong to a family of domains $\mathscr{B}$ such that an arbitrary domain G is a union of a subfamily of $\mathscr{B}$. For example, $\mathscr{B}$ might be the family of all simply connected domains, the family consisting of the interiors and exteriors of all circles (see Remark 1, p. 264), etc.

THEOREM 8.2. *Let $F(z)$ be a complete analytic function, and let M be the union of the domains of all elements of $F(z)$. Then M is itself a domain, called the domain of existence of $F(z)$.*[5]

Proof. Since the class $\mathscr{K}$ defining $F(z)$ is nonempty, so is the set M. Moreover M is open, since $z_0 \in M$ implies $z_0 \in G$ where $\{G, f\} \in \mathscr{K}$, and hence z_0 has a neighborhood $\mathscr{N}(z_0) \subset G \subset M$. To complete the proof, we now show that M is (arcwise) connected. Let z_0 and $\tilde{z}$ be two distinct points of M. Then there are elements $\{G_0, f_0\}$, $\{\tilde{G}, \tilde{f}\}$ such that $z_0 \in G_0$, $\tilde{z} \in \tilde{G}$, and also a chain

$$\{G_0, f_0\}, \{G_1, f_1\}, \ldots, \{G_n, f_n\} = \{\tilde{G}, \tilde{f}\}$$

---

[4] The term *single-valued branch* has already appeared repeatedly in this course (in special cases).

[5] This term is explained by the fact that an element of $F(z)$ is defined at a point z_0 if and only if z_0 belongs to the domain of existence of $F(z)$.

of elements joining $\{G_0, f_0\}$ and $\{\tilde{G}, \tilde{f}\}$. Since $\{G_{k-1}, f_{k-1}\}$ and $\{G_k, f_k\}$ are direct analytic continuations of each other, the set $G_{k-1} \cap G_k$ is nonempty. Let ζ_k be a point of $G_{k-1} \cap G_k$, and let γ_k be a continuous curve contained in G_k joining ζ_k to ζ_{k+1} $(k = 1, \ldots, n-1)$. Then, joining z_0 to ζ_1 by a curve $\gamma_0 \subset G_0$ and ζ_n to $\tilde{z}$ by a curve $\gamma_n \subset G_n$, we obtain a curve[6]

$$\gamma = \gamma_0 + \gamma_1 + \cdots + \gamma_n \subset M$$

joining z_0 to $\tilde{z}$, i.e., M is connected, as asserted.

DEFINITION 5. *An element $\{D, \varphi\}$ is said to be subordinate to an element $\{G, f\}$ if $D \subset G$ and $\varphi(z) = f(z)$ for all $z \in D$. This fact is expressed by writing*

$$\{D, \varphi\} \subset \{G, f\}.$$

Remark 1. If $\{D, \varphi\} \subset \{G, f\}$, then obviously $\{D, \varphi\}$ is a direct analytic continuation of $\{G, f\}$. Thus if $\{G, f\}$ belongs to a class $\mathscr{K}$ defining a complete analytic function, every element subordinate to $\{G, f\}$ also belongs to $\mathscr{K}$.

Remark 2. If $\{D, \varphi\}$ is a direct analytic continuation of $\{G, f\}$, and if $D \subset G$, then $\{D, \varphi\} \subset \{G, f\}$. In fact, there exists a domain $g \subset D \subset G$ on which $f(z)$ and $\varphi(z)$ coincide. But then $f(z)$ and $\varphi(z)$ coincide on D also, by the interior uniqueness theorem.

Example 1. Let $f(z)$ be meromorphic in the finite plane,[7] so that the singular points of $f(z)$ consist of at most countably many poles in the finite plane and a possible pole, essential singular point or limit point of poles at infinity (cf. p. II.298). Then $f(z)$ is single-valued and analytic on the domain G obtained by excluding these singular points from the extended plane.

We now show that the class $\mathscr{K}$ containing the element $\{G, f\}$ consists entirely of elements subordinate to $\{G, f\}$. Clearly, it is sufficient to show that every direct continuation of $\{G, f\}$ (or of an element subordinate to $\{G, f\}$) is subordinate to $\{G, f\}$.[8] Suppose, to the contrary, that there is an element $\{D, \varphi\}$ which is a direct continuation of $\{G, f\}$ without being subordinate to $\{G, f\}$. Then D must contain points not in G as well as points in G, and hence D must contain singular points of $f(z)$. Let z_0 be one of these points. Since $f(z)$ and $\varphi(z)$ coincide on some domain $g \subset G \subset D$, and since $G \cap D$ is itself a domain (why?), $f(z)$ and $\varphi(z)$ must coincide on $G \cap D$ itself. Therefore $f(z)$ and $\varphi(z)$ coincide in some deleted neighborhood $0 < |z - z_0| < \rho$, while at the same time z_0 is both an isolated singular point

---

[6] Cf. footnote 12, p. I.156.

[7] In particular, $f(z)$ might be entire.

[8] For simplicity, the adjective "analytic" is often omitted in the phrase "analytic continuation."

of $f(z)$ and a regular point of $\varphi(z)$ [in the sense of Sec. I.84]. But this is clearly impossible, and hence, by contradiction, $\{D, \varphi\}$ must be subordinate to $\{G, f\}$. Therefore the domain of existence of the complete analytic function $F(z)$ defined by the class $\mathscr{K}$ coincides with G, and the values of $F(z)$ coincide with those of $f(z)$ [$F(z)$ takes no other values]. As a result, we can identify $F(z)$ and $f(z)$, regarding every function meromorphic in the finite plane as a complete analytic function.

Example 2. Let G be any domain whose boundary consists of a finite number of disjoint Jordan curves. Let $\{z_n\}$ be a sequence of distinct points of G with no limit points in G, such that every boundary point of G is a limit point of $\{z_n\}$ (see Problem 8.3). Then, by Theorem 3.3, there exists a single-valued function $f(z) \not\equiv 0$ analytic on G which vanishes at every point of $\{z_n\}$. Consider the element $\{G, f\}$. As we now show, the class $\mathscr{K}$ containing $\{G, f\}$ consists entirely of elements subordinate to $\{G, f\}$.

Suppose, to the contrary, that there is an element $\{D, \varphi\}$ which is a direct continuation of $\{G, f\}$ without being subordinate to $\{G, f\}$. Then D must contain points not in G as well as points in G. Let g' be the connected component of $G \cap D$ on which $f(z)$ and $\varphi(z)$ coincide, as in the remark on p. 257. Since g' cannot coincide with D, at least one boundary point of g' must be an interior point of D (why?). Let ζ be such a point. Then ζ cannot be an interior point of G, since otherwise g' could be made larger by adjoining some neighborhood of ζ. Therefore ζ, being a limit point of G, must be a boundary point of G.

Now let $\mathscr{N}(\zeta)$ be a neighborhood of ζ which is contained in D. Then among the connected components of $G \cap \mathscr{N}(\zeta)$ there is at least one, which we denote by g'', whose boundary contains ζ and which contains infinitely many points of $\{z_n\}$ with limit point ζ (see Problem 8.4). But since $g'' \subset g'$, every point $z_n \in g''$ is a zero of both $f(z)$ and $\varphi(z)$. Therefore the point ζ, which is an interior point of D, is a limit point of zeros of $\varphi(z)$. It follows from the interior uniqueness theorem that $\varphi(z) \equiv 0$ on D, and hence $f(z) \equiv 0$ on G. This contradicts our assumption that $f(z) \not\equiv 0$ and shows that $\{D, \varphi\}$ must be subordinate to $\{G, f\}$. Hence, as in the preceding example, the domain of existence of the complete analytic function $F(z)$ defined by $\mathscr{K}$ coincides with G, and the values of $F(z)$ coincide with those of $f(z)$. As a result, we can identify $F(z)$ with the single-valued analytic function $f(z)$ defined on G. In cases like those considered in Examples 1 and 2, where the domain of definition G of a single-valued analytic function $f(z)$ coincides with its domain of existence, the boundary of G is called the *natural boundary* of $f(z)$.

Example 3. The multiple-valued function $\operatorname{Ln} z$ defined on the domain $E: 0 < |z| < \infty$ has a countably infinite set of single-valued analytic branches on every simply connected subdomain $G \subset E$. Consider all possible single-valued branches of $\operatorname{Ln} z$ defined on all possible simply connected subdomains

$G \subset E$.[9] denoting the branches generically by $f(z)$. Then, as we now show, the set $\mathscr{K}$ of all such elements $\{G, f\}$ forms a class in the sense of Definition 4 and Remark 2, p. 259. In other words,

1. Any two elements $\{G, f\}$, $\{D, \varphi\} \in \mathscr{K}$ are continuations of each other;
2. If $\{D, \varphi\}$ is a continuation of an element $\{G, f\} \in \mathscr{K}$ and if D is simply connected, then $\{D, \varphi\}$ itself belongs to $\mathscr{K}$.

To prove these assertions, we first assume that G is the domain bounded by the ray $0 \leqslant x < \infty$, $y = 0$. Then the single-valued branches of Ln z on G are all of the form

$$f_{2k} = \ln |z| + i\theta_{2k} \qquad (k = 0, \pm 1, \pm 2, \ldots),$$

where θ_{2k} is the value of Arg z satisfying the condition

$$2k\pi < \theta_{2k} < 2(k + 1)\pi. \qquad (8.4)$$

Then obviously $\{G, f_{2m}\}$ and $\{G, f_{2n}\}$ are continuations of each other for arbitrary m and n (just as in the example on pp. 258–259). Next suppose $D \subset E$ is an arbitrary simply connected domain, and let $\varphi(z)$ be any single-valued branch of Ln z defined on D. Let z_0 be a point of D which does not belong to the boundary of G (so that z_0 is not a nonnegative real number). Then z_0 has a neighborhood $\mathscr{N}(z_0)$ containing no boundary points of G, and in this neighborhood $\varphi(z)$ is of the form

$$\varphi(z) = \ln |z| + i\theta_{2k_0},$$

since for $z \in \mathscr{N}(z_0)$, every value of Arg z satisfies a condition of the form (8.4). Therefore $\varphi(z)$ coincides with $f_{2k_0}(z)$ on $\mathscr{N}(z_0)$, i.e., $\{D, \varphi\}$ is a (direct) continuation of $\{G, f_{2k_0}\}$. This proves the first assertion.

To prove the second assertion, it is sufficient to consider the case of *direct* continuation. Thus suppose $\{D, \varphi\}$ (where D is simply connected) is a direct continuation of an element $\{G, f\} \in \mathscr{K}$, so that $\varphi(z)$ coincides with $f(z)$ on some domain $g \subset G \cap D$. Then $\{D, \varphi'\}$ is clearly a direct continuation of $\{G, f'\}$, where the prime denotes differentiation with respect to z. But

$$f'(z) = \frac{1}{z} \qquad (z \in G),$$

and hence

$$\varphi'(z) = \frac{1}{z} \qquad (z \in g),$$

which implies

$$\varphi'(z) = \frac{1}{z} \qquad (z \in D). \qquad (8.5)$$

---

[9] Of course, single-valued branches of Ln z can also be defined on appropriate multiply connected domains.

Therefore D cannot contain the point $z = 0$, since otherwise $\varphi(z)$ would not be analytic on D. Moreover, D cannot contain the point $z = \infty$, since then $\varphi(z)$, being analytic at ∞, would have a Laurent expansion at ∞ of the form

$$\varphi(z) = a_0 + \frac{a_1}{z} + \frac{a_2}{z^2} + \cdots,$$

and hence

$$\varphi'(z) = -\frac{a_1}{z^2} - \frac{2a_2}{z^3} - \cdots,$$

which is incompatible with (8.5). It follows that the simply connected domain D is contained in $E{:}0 < |z| < \infty$.

Now let z_0 be a point of $G \cap D$ such that $f(z)$ and $\varphi(z)$ coincide in a neighborhood $\mathcal{N}(z_0)$, and let $\varphi_0(z)$ be any single-valued branch of Ln z defined on D. Since

$$\varphi(z) - \varphi_0(z) = f(z) - \varphi_0(z) = 2k_0\pi i$$

on $\mathcal{N}(z_0)$,[10] we must have

$$\varphi(z) - \varphi_0(z) = 2k_0\pi i$$

on the whole domain D. Therefore

$$\varphi(z) = \varphi_0(z) + 2k_0\pi i$$

is also a single-valued branch of Ln z on D, and the second assertion is proved.

Thus we have shown that the class $\mathcal{K}$ of all single-valued branches of Ln z defined on all simply connected domains contained in $0 < |z| < \infty$ defines a complete analytic function, in the sense of Definition 4 and Remark 2, p. 259. This allows us to regard Ln z as a complete analytic function, with domain of existence $0 < |z| < \infty$.

40. Circular Elements. The Monodromy Theorem

The case where G is a domain of the form $|z - z_0| < r$ (the interior of a circle) or of the form $|z| > r$ (the exterior of a circle[11]) is of particular importance. Then the element $\{G, f\}$ is called a *circular element*,[12] and the function $f(z)$ analytic on G can be represented as the sum of a convergent power series

$$f(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n \qquad (|z - z_0| < r) \tag{8.5}$$

---

[10] Bear in mind that $f(z)$ and $\varphi_0(z)$ are both single-valued branches of Ln z on $\mathcal{N}(z_0)$.

[11] There is no need to consider domains of the form $|z - z_0| > r$ (recall Problem II.1.4).

[12] Already encountered in Sec. I.84 for the case $|z - z_0| < r$.

or of a convergent Laurent series

$$f(z) = \sum_{n=0}^{\infty} a_n z^{-n} \qquad (|z| > r) \tag{8.6}$$

containing no positive powers of z.[13] Thus a circular element e is uniquely characterized by a series of the form (8.5) or (8.6) and by the circle whose interior or exterior serves as the domain of the element. Such an element will often be denoted explicitly by

$$e\colon \sum_{n=0}^{\infty} a_n(z - z_0)^n \qquad (|z - z_0| < r), \tag{8.7}$$

or by

$$e\colon \sum_{n=0}^{\infty} a_n z^{-n} \qquad (|z| > r), \tag{8.8}$$

rather than by the more concise notation $e = \{K, f(z)\}$, where K is the set $|z - z_0| < r$ or $|z| > r$.

DEFINITION 1. *A circular element of the form* (8.7) *is said to have series*

$$\sum_{n=0}^{\infty} a_n(z - z_0)^n,$$

center z_0, *radius* r *and disk* $|z - z_0| < r$. *A circular element of the form* (8.8) *is said to have series*

$$\sum_{n=0}^{\infty} a_n z^{-n},$$

center ∞, *radius* r *and "disk"* $|z| > r$.

Remark 1. As already noted (recall Remark 2, p. 259), a complete analytic function $F(z)$ can be defined just as well by a class $\mathscr{K}_0$ of circular elements,[14] as by a class $\mathscr{K} \supset \mathscr{K}_0$ of arbitrary elements. This means more than the abstract statement that knowledge of $\mathscr{K}_0$ allows us to decide whether or not a given noncircular element $\{G, f\}$ belongs to $\mathscr{K} \supset \mathscr{K}_0$. In fact, an arbitrary element $\{G, f\} \in \mathscr{K}$ has a series expansion in a neighborhood $\mathscr{N}(z_0) \subset G$ of every point $z_0 \in G$, and this neighborhood and series define a circular element subordinate to $\{G, f\}$.

---

[13] The absence of positive powers guarantees that $f(z)$ is analytic at $z = \infty$ (where it has the value a_0).

[14] Two circular elements e and e' belong to $\mathscr{K}_0$ if and only if they are continuations of each other. In general, the chain of direct continuations joining e and e' contains noncircular elements, but this can always be avoided (see Problem 8.6).

Remark 2. Given a circular element (8.7), let ζ be any point of the disk $|z - z_0| < r$, and let

$$\rho = \min \{|\zeta - z_0|, r - |\zeta - z_0|\},$$

so that the series

$$\sum_{k=0}^{\infty} a_n'(z - \zeta)^n, \tag{8.9}$$

where

$$a_n' = \frac{f^{(n)}(\zeta)}{n!},$$

converges for $|z - \zeta| < \rho$. Then the element

$$e' : \sum_{n=0}^{\infty} a_n'(z - \zeta)^n \qquad (|z - \zeta| < \rho)$$

is subordinate to the element (symbolically $e' \subset e$). We shall also say that the series (8.9) is subordinate to the series

$$\sum_{n=0}^{\infty} a_n(z - z_0)^n.$$

DEFINITION 2. *If L is a (continuous) curve with equation*

$$z = \lambda(t) \qquad (\alpha \leqslant t \leqslant \beta),$$

and if $\alpha \leqslant \alpha' < \beta' \leqslant \beta$, then $L[\alpha', \beta']$ denotes the arc of L with equation

$$z = \lambda(t) \qquad (\alpha' \leqslant t \leqslant \beta').$$

Similarly, $L[\alpha', \beta')$ denotes the "half-open" arc

$$z = \lambda(t) \qquad (\alpha' \leqslant t < \beta').$$

DEFINITION 3. *Let*

$$L : z = \lambda(t) \qquad (\alpha \leqslant t \leqslant \beta)$$

be a curve and let $e = \{K, f(z)\}$ be a circular element with its center at the initial point $z_0 = \lambda(\alpha)$ of L. Suppose there exists a finite set of circular elements

$$e_j = \{K_j, f_j(z)\} \qquad (j = 0, 1, \ldots, n)$$

such that

1. *Each disk K_j has its center at a point $z_j = \lambda(t_j) \in L$, where*[15]

$$\alpha = t_0 \leqslant t_1 \leqslant \cdots \leqslant t_n = \beta;$$

2. *$L_j = L[t_{j-1}, t_j] \subset K_{j-1} \cup K_j$ for every $j = 1, \ldots, n$;*[16]

[15] In writing $\leqslant$ instead of $<$, we allow for the possibility of some of the disks K_j having the same centers.

[16] If $t_{j-1} = t_j$, the $L_j = \{t_j\}$ and condition 3 is obvious.

3. *$e_0 = e$ and e_j is a direct continuation of e_{j-1} for every $j = 1, \ldots, n$. Then the element e is said to be analytically continued along L (by the chain $e = e_0, e_1, \ldots, e_n$), and the element e_n is called the result of continuing e along L. The chain $e = e_0, e_1, \ldots, e_n$ is also said to join e to e_n along L.*

Remark 1. Analytic continuation along an arc $L[\alpha, \tau] \subset L$ is defined in the obvious way, with $L[\alpha, \tau]$ playing the role of L and τ that of β.

Remark 2. Suppose a chain $e_0, e_1, \ldots, e_{n-1}, e_n$ continues the element e_0 along the curve L. Then, according to Definition 3, the chain $e_n, e_{n-1}, \ldots, e_1, e_0$ continues e_n along the arc $-L$ (the arc L traversed in the opposite direction), with the original element e_0 as the result of the continuation.

Remark 3. An arbitrary continuation can be regarded as a continuation along some curve. In fact, let $e_0, e_1, \ldots, e_n$ be a chain of elements joining e_0 to e_n,[17] and draw the line segments joining the center of e_{j-1} to the center of e_j for every $j = 1, \ldots, n$, thereby obtaining a polygonal curve Λ. Then the chain $e_0, e_1, \ldots, e_n$ joins e_0 to e_n along Λ.

Definition 3 tacitly relies upon

THEOREM 8.3. *The result of analytic continuation of an element e along a curve L is independent of the chain of elements continuing e. More precisely, if $e_0, e_1, \ldots, e_n$ and $e'_0, e'_1, \ldots, e'_{n'}$ are two distinct chains of elements continuing $e = e_0 = e'_0$ along a curve*

$$L: z = \lambda(t) \qquad (\alpha \leqslant t \leqslant \beta),$$

then one of the elements e_n and $e'_{n'}$ is subordinate to the other.

Proof. Let

$$e_j = \{K_j, f_j(z)\}, \qquad e'_l = \{K'_l, \varphi_l(z)\},$$

where

$$z_j = \lambda(t_j), \quad z'_l = \lambda(t'_l) \qquad (j = 0, 1, \ldots,; l = 0, 1, \ldots, n').$$

Each chain defines a single-valued analytic function on L, provided that a point $z' = \lambda(t') \in L$ is regarded as distinct from a point $z'' = \lambda(t'') \in L$ if $t' \neq t''$ (just as in the remark on p. I.60).[18] In fact, if $z = \lambda(t)$ is a point of the arc

$$L_j = L[t_{j-1}, t_j] \subset K_{j-1} \cup K_j,$$

---

[17] Henceforth, unless the contrary is explicitly stated, an "element" will mean a "circular element" and a "continuation" will involve only circular elements.

[18] Otherwise, the same point $z \in L$ can correspond to different values of the parameter t, and hence to different arcs L_j and different pairs of disks K_{j-1}, K_j. In general, this makes the functions $f(z)$ and $\varphi(z)$ multiple-valued.

we write

$$f(z) = \begin{cases} f_{j-1}(z) & \text{if } z \in K_{j-1}, \quad z \notin K_j, \\ f_j(z) & \text{if } z \notin K_{j-1}, \quad z \in K_j, \\ f_{j-1}(z) = f_j(z) & \text{if } z \in K_{j-1} \cap K_j \end{cases}$$

(recall that e_j is a direct continuation of e_{j-1}). Similarly, the same considerations applied to the chain $e_0', e_1', \ldots, e_{n'}'$ lead to another single-valued analytic function $\varphi(z)$ defined on L.

The functions $f(z)$ and $\varphi(z)$ coincide on $L \cap K_0$, since K_0 is the disk of both e_0 and e_0'. Let $T \in (\alpha, \beta]$ be the least upper bound of the values of t' such that $f[\lambda(t)] = \varphi[\lambda(t)]$ for all $t \in [\alpha', t']$. If $T < \beta$, there exist indices m and m' such that

$$t_{m-1} \leqslant T \leqslant t_m, \qquad t'_{m'-1} \leqslant T \leqslant t'_{m'}.$$

Therefore $Z = \lambda(T)$ belongs to both $K_{m-1} \cup K_m$ and $K'_{m'-1} \cup K'_{m'}$. To be explicit, suppose $Z \in K_{m-1} \cap K'_{m'}$. Then $f(z)$ and $\varphi(z)$ are both single-valued and analytic in some neighborhood $\mathcal{N}(Z) \subset K_{m-1} \cap K'_{m'}$ (why?). But $f(z)$ and $\varphi(z)$ coincide on the set $L[\alpha, \tau] \subset \mathcal{N}(Z)$, and hence, by the interior uniqueness theorem, they coincide on $\mathcal{N}(Z)$ itself, contrary to the definition of T. This contradiction shows that $T = \beta$. Then, if $\zeta = \lambda(\beta)$, the same argument shows that $f(z)$ and $\varphi(z)$ coincide on $K_n \cap K'_{n'}$, i.e., one of the elements e_n and $e_{n'}'$ is subordinate to the other, as asserted.

Next we establish an important

LEMMA. *Given an element* $e_0 = \{K, f(z)\}$, *let*

$$L: z = \lambda(t) \qquad (\alpha \leqslant t \leqslant \beta)$$

be a curve contained in the disk K, *and let* ε *be an element subordinate to* e_0, *with disk* K_0 *centered at the initial point* $z_0 = \lambda(\alpha)$ *of* L. *Then* ε *can be continued along* L, *and the resulting element is subordinate to* e_0 (*if its disk is sufficiently small*).

Proof. Let $\rho > 0$ be the distance between L and the boundary of K (choose $\rho = 1$ if this distance is infinite). Using the uniform continuity of $z = \lambda(t)$ on L, we find a partition

$$\alpha = t_1 < t_2 < \cdots < t_n = \beta$$

of the interval $[\alpha, \beta]$ into subintervals so small that

$$L_j = L[t_{j-1}, t_j] \subset K_{j-1} \qquad (j = 1, \ldots, n),$$

where K_j is the disk $|z - z_j| < \rho$ centered at the point $z_j = \lambda(t_j)$. Then the disks $K_0, K_1, \ldots, K_n$ satisfy the first two properties of Definition 3, while the elements

$$\varepsilon_j = \{K_j, f(z)\} \qquad (j = 0, 1, \ldots, n)$$

obviously satisfy the third property. (Note that every ε_j is subordinate to e_0.) Therefore $\varepsilon = \varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n$ is a chain joining ε to ε_n along L. According to Theorem 8.3, the result of any other continuation of ε along L is subordinate to ε_n and hence to e_0, if its disk is sufficiently small.

Theorem 8.3 shows that the result of continuing an element e along a curve L is independent of the chain of elements continuing e. We now show that, to a certain extent, the result is also independent of L:

THEOREM 8.4. *Let $e_0, e_1, \ldots, e_n$ be a chain of elements, with disks $K_0, K_1, \ldots, K_n$, joining e_0 to e_n along a curve*

$$L: z = \lambda(t) \qquad (\alpha \leqslant t \leqslant \beta).$$

Suppose there is another curve

$$L': z = \mu(t') \qquad (\alpha' \leqslant t' \leqslant \beta')$$

with the same end points

$$z_0 = \lambda(\alpha) = \mu(\alpha'), \qquad z' = \lambda(\beta) = \mu(\beta'),$$

and a corresponding partition

$$\alpha' = t_0' \leqslant t_1' \leqslant \cdots \leqslant t_n' \leqslant t_{n+1}' = \beta'$$

of the interval $[\alpha', \beta']$ such that

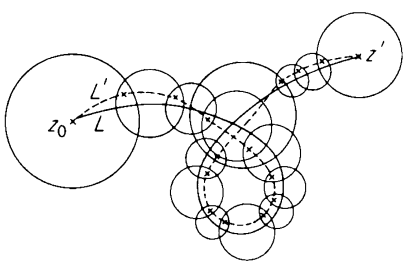


FIGURE 8.1

1. *Except for z_0 and z', every point $z_j' = \mu(t_j')$ is contained in $K_{j-1} \cap K_j$;*
2. *Every arc*

   $$L_j' = L'[t_{j-1}', t_j'] \qquad (j = 1, \ldots, n+1)$$

   is contained in K_{j-1}.[19]

Then e_0 can be continued along L', with the same result as obtained by continuing e_0 along L.

Proof. Using the lemma, we continue $e_0 = \varepsilon_0^{(0)}$ along $L_0' \subset e_0$, making sure that the disk of the final element $\varepsilon_{k_0}^{(0)}$ (with center z_1') is contained in $K_0 \cap K_1$.[20] Then $\varepsilon_{k_0}^{(0)}$ is subordinate to both e_0 and e_1 (why?). Therefore, using the lemma again, this time applied to the elements e_1, $\varepsilon_{k_0}^{(0)} \subset e_1$ and the curve L_1', we continue $\varepsilon_{k_0}^{(0)} = \varepsilon_0^{(0)}$ along L_1', making sure that the disk

[19] The geometric meaning of these conditions is shown in Figure 8.1, for the case where the disks $K_0, K_1, \ldots, K_n$ have distinct centers.

[20] If $e_0 = \{K_0, f_0\}$, this can always be achieved by adding an element of the form $\{K, f_0\}$, where $K \subset K_0 \cap K_1$, to the end of the chain constructed in the proof of the lemma.

of the final element $\varepsilon_{k_1}^{(1)}$ (with center z_2') is contained in $K_1 \cap K_2$. The final element $\varepsilon_{k_1}^{(1)}$ is then subordinate to both e_1 and e_2, and the chain

$$e_0 = \varepsilon_0^{(0)}, \ldots, \varepsilon_{k_0}^{(0)} = \varepsilon_0^{(1)}, \ldots, \varepsilon_{k_1}^{(1)}$$

joins e_0 to $\varepsilon_{k_1}^{(1)}$ along $L'[\alpha', t_2'] \subset L'$. Repeating the same argument $n - 1$ more times, and choosing e_n itself as the final element (why is this possible?), we finally obtain a chain

$$e_0 = \varepsilon_0^{(0)}, \ldots, \varepsilon_{k_0}^{(0)} = \varepsilon_0^{(1)}, \ldots, \varepsilon_{k_1}^{(1)} = \varepsilon_0^{(2)}, \ldots, \varepsilon_{k_{n-1}}^{(n-1)} = \varepsilon_0^{(n)}, \ldots, \varepsilon_{k_n}^{(n)} = e_n$$

joining e_0 to e_n along $L' = L'[\alpha', t_{n+1}']$. Since e_n is also the result of continuing e_0 along L, the proof is now complete.

COROLLARY. *Given an element e_0 with center z_0, let L be the line segment joining z_0 to a point $z' \neq z_0$. Then z' has a neighborhood $\mathcal{N}(z')$ such that e_0 can be continued along any polygonal curve of the form*[21]

$$L' = \overline{z_0\zeta} + \overline{\zeta z'}, \qquad \zeta \in \mathcal{N}(z'),$$

with the same result as obtained by continuing e_0 along L.

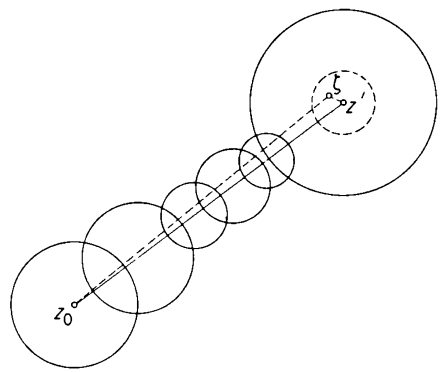


FIGURE 8.2

Proof. The idea of the proof is immediately apparent from Figure 8.2.

We are now in a position to prove one of the key propositions of analytic function theory:

THEOREM 8.5 (*Monodromy theorem*). *Given a simply connected domain G, let e_0 be an element with disk $K \subset G$ centered at z_0. Suppose e_0 can be continued along every continuous curve in G, thereby defining a function $f(z)$ at every point of G.*[22] *Then $f(z)$ is single-valued on G.*

Proof. Suppose, to the contrary, that $f(z)$ is multiple-valued on G, and hence takes at least two distinct values at some point $z_1 \in G$. Then we can find (at least) two distinct elements e_1' and e_1'', with disks in G and

---

[21] By $\overline{z_0\zeta}$ is meant the line segment joining z_0 to ζ, and similarly for $\overline{\zeta z'}$. L' is the polygonal curve going from z_0 to ζ along $\overline{z_0\zeta}$ and then from ζ to z' along $\overline{\zeta z'}$.

[22] More explicitly, given any $z_1 \in G$, $z_1 \neq z_0$, let $L_0 \subset G$ be a curve joining z_0 to z_1, and let e_1 be the result of continuing e_0 along L, with disk $K_1 \subset G$ and series

$$\sum_{n=0}^{\infty} a_n(z - z_1)^n \qquad (z \in K_1).$$

Then $f(z_1) = a_0$ by definition.

centers at z_1, and two distinct curves $L_1, L_2 \subset G$ joining z_0 to z_1 such that e_1' is the result of continuing e_0 along L_1, while e_1'' is the result of continuing e_0 along L_2. Therefore continuation of e_1' along $-L_2$ must give an element e_0' distinct from e_0,[23] since otherwise continuation of e_0 along L_2 would give e_1' instead of e_1'' (recall Remark 2, p. 266). It follows that continuation of e_0 along the closed curve $L = L_1 - L_2 \subset G$, going from z_0 to z_1 along L_1 and returning to z_0 along L_2, gives an element e_0' distinct from e_0. Moreover, this assertion remains true if we replace L by a suitable polygonal curve contained in G. In fact, let $e_0, e_1, \ldots, e_n = e_0'$ be the chain joining e_0 to e_0' along L, and replace the elements e_j by subordinate elements ε_j so small that the arcs of L joining centers of consecutive ε_j are all shorter than the distance from L to the boundary of G. We can then replace L by the polygonal curve $\Lambda \subset G$ made up of the line segments joining centers of consecutive ε_j without changing the result of the continuation, i.e., continuation of e_0 along Λ still leads to the element e_0' distinct from e_0.

Next, by considerations resembling those used to prove Cauchy's integral theorem (see Sec. I.65),[24] we replace Λ first by a closed polygonal *Jordan* curve (with no self-intersections) contained in G, and finally by a triangular contour contained in G. The argument goes as follows: Starting from the point z_0 and moving along the curve Λ, we eventually arrive at points which have already been encountered. Let z_1 be the first such point. If $z_1 = z_0$, then Λ has no self-intersections from the outset, and we can proceed to the next step of the construction. If $z_1 \neq z_0$, let Λ^* be the closed polygonal curve obtained by starting from z_1 and traversing Λ until the first return to the point z_1. Obviously, Λ^* has no self-intersections. Let e_1 be the result of continuing the element e_0 along Λ from z_0 to the first encounter with the point z_1. If continuation of e_1 along Λ^* leads to an element e_1' distinct from e_1, we replace Λ by Λ^* in our subsequent considerations. Otherwise we remove Λ^* from Λ, thereby obtaining a new polygonal curve $\Lambda_0 \subset G$, with one self-intersection less than Λ, such that continuation of e_0 along Λ_0 leads to an element e_0' distinct from e_0. Repeating this operation of removing closed polygonal curves a finite number of times, we finally obtain a closed polygonal Jordan curve $\Lambda' \subset G$ such that the result of continuing some element e' along Λ' is an element distinct from e' (of course, e' itself is the result of continuing e_0 along a curve contained in G). In this process of eliminating self-intersections from Λ, we may encounter segments traversed twice in opposite directions, but these can be removed at once since an element does not change as a result of continuation back and forth along the same segment.

---

[23] Here, of course, we mean "distinct even if defined on the same disk."

[24] Also recall Theorem I.4.12.

We now divide the polygon $\overline{I(L')}$ bounded by L' into convex polygons (as in Step 4, p. I.265), and afterwards divide each of the resulting polygons into triangles (as in Step 3, p. I.264).[25] Let $\delta = \overline{\zeta_0\zeta'}$ be a side of one of these triangles such that the only points of δ belonging to Λ' are its end points ζ_0 and ζ', where ζ_0 precedes ζ' as Λ' is traversed in the positive direction.[26] Let ε_0 be the result of continuing the element e' along Λ' from the initial point z' to the point ζ_0. Moreover, let ε' be the result of continuing ε_0 along δ, and let ε'' be the result of continuing ε_0 along the part of Λ' joining ζ_0 to ζ'. Then the elements ε' and ε'' may coincide, in the sense that one is subordinate to the other. In this case, we replace the part of Λ' joining ζ_0 to ζ' by the segment δ, thereby obtaining a new closed polygonal (Jordan) curve Λ'', forming the boundary of a union of fewer triangles than Λ', such that the result of continuing e' along Λ'' is an element distinct from e'. On the other hand, if ε'' does not coincide with ε', we replace Λ' by the closed polygonal curve $\tilde{\Lambda}$ consisting of the part of Λ' going from ζ_0 to ζ', together with the segment $-\delta = \overline{\zeta'\zeta_0}$. Then the result of continuing ε_0 along $\tilde{\Lambda}$ is an element distinct from ε_0. Repeating this argument, we successively replace Λ' by a curve bounding fewer and fewer triangles, until finally, after a finite number of steps, we arrive at a triangle $\Delta \subset G$ and an element ε (itself the result of continuing e_0 along a curve contained in G), whose disk is centered at a vertex ζ of Δ, such that the result of continuing ε along Δ is distinct from ε.

Drawing the line segment from ζ to the midpoint of the side of Δ opposite ζ, we divide Δ into two triangles, at least one of which, say Δ_1, has the same property as Δ, i.e., the result of continuing ε along Δ_1 is an element distinct from ε (why?). Then, joining ζ to the midpoint of the side of Δ_1 opposite ζ, we obtain a new triangle Δ_2, with the same property as Δ and Δ_1. Repeating this process indefinitely, we construct a "nested" sequence of triangles

$$\Delta = \Delta_0 \supset \Delta_1 \supset \Delta_2 \supset \cdots \supset \Delta_n \supset \cdots,$$

all with a common vertex ζ, such that

$$\delta = \delta_0 \supset \delta_1 \supset \delta_2 \supset \cdots \supset \delta_n \supset \cdots,$$

where δ_n denotes the side of Δ_n opposite ζ. Moreover, the length of δ_n clearly approaches zero as $n \to \infty$. Hence there is a unique point η contained in all the δ_n (cf. p. I.29). But, according to the corollary to Theorem 8.4, η has a neighborhood $\mathcal{N}(\eta)$ such that the result of continuing ε along any polygonal curve

$$\overline{\zeta z} + \overline{z\eta}, \qquad z \in \mathcal{N}(\eta)$$

---

[25] For simplicity, we use the same word "triangle" for both triangular contours and the closed domains bounded by such contours (cf. footnote 3, p. I.262).

[26] I.e., the direction in which Λ' is traversed in the original curve Λ.

is the same as the element ε obtained by continuing ε along the segment $\overline{\zeta\eta}$. Let η_n' and η_n'' denote the end points of δ_n. Then $\eta_n', \eta_n'' \in \mathcal{N}(\eta)$ for all sufficiently large n, and hence the result of continuing ε along Δ_n is the same as that of first continuing ε along $\overline{\zeta\eta_n'} + \overline{\eta_n'\eta}$, which gives ε, and then continuing ε along $\overline{\eta\eta_n''} + \overline{\eta_n''\zeta}$, which gives back ε, contrary to the construction of Δ_n. Therefore the initial assumption that $f(z)$ is multiple-valued on G must be false, and the theorem is established by contradiction.

COROLLARY. *The function $f(z)$ figuring in Theorem* 8.5 *is analytic on G.*

Proof. Continuing e_0 along a curve $L_0 \subset G$ joining z_0 to a distinct point $z_1 \in G$, we obtain an element e_1, with disk $K_1 \subset G$ centered at z_1 and series

$$\sum_{n=0}^{\infty} a_n(z - z_1)^n \qquad (z \in K_1) \tag{8.10}$$

(recall footnote 22, p. 269). We want to show that (8.10) represents $f(z)$ in K_1. To see this, let $L_1 \subset G$ be a curve joining z_0 to a point $\zeta \in K_1$ and $L_2 \subset G$ a curve joining ζ to z_1, where $L_1 + L_2 \neq L_0$. Then, since the result of continuing e_0 along $L_1 + L_2$ is the same as the element e_1 obtained by continuing e_0 along L_0, the result of continuing e_1 along $-L_2$ must be the same as the element

$$\varepsilon\colon \sum_{n=0}^{\infty} b_n(z - \zeta)^n \qquad (|z - \zeta| < \rho)$$

obtained by continuing e_0 along L_1 (why?). According to the lemma on p. 267, ε is subordinate to e_1 for sufficiently small ρ. But then

$$f(\zeta) = b_0 = \sum_{n=0}^{\infty} a_n(\zeta - z_1)^n \qquad (\zeta \in K_1),$$

i.e., $f(z)$ is analytic in a neighborhood of z_1. Therefore $f(z)$ is analytic on G, since $z_1 \in G$ is arbitrary (how about the case $z_1 = z_0$?).

41. Analytic Continuation in a Star

Next we consider the problem of continuing a given element e_0 with center z_0 along various rays (half-lines) emanating from z_0. Obviously, every ray has an initial portion, of length no less than the radius of e_0, along which continuation is possible. If an arbitrary point of a given ray L cannot be reached by continuing e_0 along L, then L must contain a point $z_1 \neq z_0$ such that e_0 can be continued along the segment $\overline{z_0\zeta}$, where ζ is any point of $\overline{z_0z_1}$ distinct from its end points, but not along the whole segment $\overline{z_0z_1}$. Suppose

we mark z_1 on L (setting $z_1 = \infty$ if e_0 can be continued along the whole ray), and do the same thing for every other ray emanating from z_0. Then the set S_{e_0} of all half-open segments of the form $\overline{z_0 z_1} - \{z_1\}$ is called the (*rectilinear*) *star* of the element e_0.[27] A set of this type, consisting of all points of the plane which can be reached by continuing e_0 along an arbitrary ray emanating from z_0, is shown schematically in Figure 8.3.

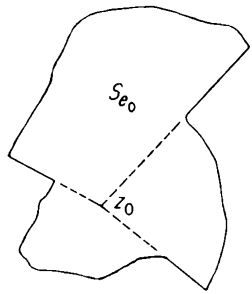


FIGURE 8.3

THEOREM 8.6. *The star S_{e_0} of an element e_0 is a simply connected domain.*

Proof. If $z_1 \neq z_0$, $z_1 \in S_{e_0}$, then, by the corollary to Theorem 8.4, z_1 has a neighborhood $\mathcal{N}(z_1)$ such that e_0 can be continued along any segment $\overline{z_0\zeta}$, $\zeta \in \mathcal{N}(z_1)$, i.e., $\mathcal{N}(z_1) \subset S_{e_0}$, and hence z_1 is an interior point of S_{e_0}. Moreover, z_0 itself is obviously an interior point of S_{e_0}, since S_{e_0} contains the whole disk of the element e_0 with center z_0. Therefore S_{e_0} is an open set. Since any two points $z_1, z_2 \in S_{e_0}$ can be joined by a polygonal curve $\overline{z_1 z_0} + \overline{z_0 z_2} \subset S_{e_0}$, the set S_{e_0} is connected as well as open, and hence is a domain. To verify that S_{e_0} is *simply* connected, suppose S_{e_0} contains a closed Jordan curve γ whose interior $I(\gamma)$ contains a point $\zeta \notin S_{e_0}$. Then the ray $\overline{z_0\zeta}$ intersects γ in a point z_1 such that $\zeta \in \overline{z_0 z_1}$. But $\overline{z_0 z_1} \subset S_{e_0}$, by the very definition of S_{e_0}, and hence $\zeta \in S_{e_0}$. This contradiction shows that if S_{e_0} contains a closed Jordan curve γ, then S_{e_0} also contains $I(\gamma)$, i.e., S_{e_0} is simply connected (by Theorem I.4.14).

Since S_{e_0} is a simply connected domain, it follows from the monodromy theorem and its corollary that the function $f(z)$ obtained by continuing e_0 along every line segment $\overline{z_0 z} \subset S_{e_0}$ (or for that matter, along every continuous curve $\widehat{z_0 z} \subset S_{e_0}$) is single-valued and analytic. We now turn to the problem of representing $f(z)$ in S_{e_0}:

THEOREM 8.7. *Given an element e_0 with center z_0 and star S_{e_0}, let $f(z)$ be the function defined by continuing e_0 along every segment $\overline{z_0 z} \subset S_{e_0}$. Then*

$$f(z) = \lim_{n\to\infty} f_n(z) = \lim_{n\to\infty} \left[c_0^{(n)} f(z_0) + c_1^{(n)} \frac{f'(z_0)}{1!}(z - z_0) + \cdots + c_{k_n}^{(n)} \frac{f^{(k_n)}(z_0)}{k_n!}(z - z_0)^{k_n} \right], \quad (8.11)$$

[27] Synonymously, the *principal star* or *Mittag-Leffler star* of e_0.

where the sequence $\{f_n(z)\}$ *converges uniformly inside* S_{e_0} *and the coefficients* $c_0^{(n)}, c_1^{(n)}, \ldots, c_{k_n}^{(n)}$ *are independent of* $f(z)$.

Proof. Let F_0 be any compact subset of S_{e_0}, and let F be the union of all line segments $\overline{z_0 z}$, $z \in F_0$. Then F is itself a compact subset of S_{e_0} (why?). Let L be a closed rectifiable Jordan curve such that $L \subset S_{e_0}$, $I(L) \supset F$ (see Problem 8.7). It follows from Cauchy's integral formula that

$$f(z) = \frac{1}{2\pi i}\int_L \frac{f(\zeta)}{\zeta - z}\,d\zeta = \frac{1}{2\pi i}\int_L \frac{1}{1 - \dfrac{z - z_0}{\zeta - z_0}}\,\frac{f(\zeta)}{\zeta - z_0}\,d\zeta \tag{8.12}$$

for all $z \in F$. Let $E(F, L)$ be the set of all points

$$w = \frac{z - z_0}{\zeta - z_0} \tag{8.13}$$

such that $z \in F$, $\zeta \in L$. Then $E(F, L)$ is compact. In fact, let $\delta > 0$ be the distance between the point z_0 and the curve L, while

$$M = \sup_{z \in F} |z - z_0|.$$

Then

$$|w| \leqslant \frac{M}{\delta}$$

for all $w \in E(F, L)$, i.e., $E(F, L)$ is bounded. Moreover, if w' is a limit point of a sequence

$$\{w_n\} = \left\{\frac{z_n - z_0}{\zeta_n - z_0}\right\},$$

we can always go over to subsequences such that $\{z_n\}$ and $\{\zeta_n\}$ converge. Since the sets F and L are closed,

$$\lim_{n\to\infty} z_n = z' \in F, \qquad \lim_{n\to\infty} \zeta_n = \zeta' \in L,$$

and hence

$$w' = \lim_{n\to\infty} w_n = \lim_{n\to\infty} \frac{z_n - z_0}{\zeta_n - z_0} = \frac{z' - z_0}{\zeta' - z_0} \in E(F, L),$$

i.e., $E(F, L)$ is closed.

Next we observe that no real number $u \geqslant 1$ can belong to $E(F, L)$. In fact, (8.13) can take positive real values only if z and ζ lie on the same ray emanating from z_0. But ζ cannot lie between z_0 and z, since $\zeta \notin F$, and hence

$$w = \frac{z - z_0}{\zeta - z_0} < 1.$$

Therefore $E(F, L)$ is contained in the domain G bounded by the segment $u \geqslant 1$, $v = 0$ (the part of the real axis going from 1 to ∞). The function

$$\frac{1}{1 - w}$$

is single-valued and analytic on G, and hence, by Runge's theorem (Theorem 3.5), there exists a sequence of polynomials

$$P_n(w) = c_0^{(n)} + c_1^{(n)} w + \cdots + c_{k_n}^{(n)} w^{k_n} \tag{8.14}$$

such that

$$\lim_{n \to \infty} P_n(w) = \frac{1}{1 - w},$$

where the convergence is uniform inside G, in particular on the compact set $E(F, L) \subset G$. Therefore

$$\lim_{n \to \infty} P_n\left(\frac{z - z_0}{\zeta - z_0}\right) = \frac{1}{1 - \dfrac{z - z_0}{\zeta - z_0}},$$

where the convergence is uniform for all $z \in E$, $\zeta \in L$. Combining this result with (8.12) and using the uniform convergence, we find that

$$f(z) = \lim_{n \to \infty} \frac{1}{2\pi i} \int_L P_n\left(\frac{z - z_0}{\zeta - z_0}\right) \frac{f(\zeta)}{\zeta - z_0}\, d\zeta, \tag{8.15}$$

where the convergence is now uniform on F (why?). Substituting (8.14) into (8.15), we obtain

$$\begin{aligned} f(z) = \lim_{n \to \infty} \Bigg[& \frac{c_0^{(n)}}{2\pi i} \int_L \frac{f(\zeta)}{\zeta - z_0}\, d\zeta + \frac{c_1^{(n)}}{2\pi i} (z - z_0) \int_L \frac{f(\zeta)}{(\zeta - z_0)^2}\, d\zeta \\ & + \cdots + \frac{c_{k_n}^{(n)}}{2\pi i} (z - z_0)^{k_n} \int_L \frac{f(\zeta)}{(\zeta - z_0)^{k_n + 1}}\, d\zeta \Bigg], \end{aligned} \tag{8.16}$$

where the convergence is uniform on $F_0 \subset F$ and hence inside S_{e_0} (since F_0 is an arbitrary compact subset of S_{e_0}). Since (8.16) is another way of writing (8.11) [recall formula (I.14.20)], the proof is now complete.

Remark. The representation (8.11) is called *Mittag-Leffler's expansion* and explicitly solves the problem of analytic continuation of a given element e_0 in its star, i.e., along all rays emanating from the center of e_0. We again call attention to the fact that the coefficients $c_0^{(n)}, c_1^{(n)}, \ldots, c_{k_n}^{(n)}$ are "universal" and can be calculated once and for all. In fact, e_0 enters formula (8.11) only through the quantities

$$\frac{f^{(k)}(z_0)}{k!},$$

which are, of course, just the coefficients of the series defining the element e_0.

42. Singular Points. Generalized Elements and the Analytic Configuration

The concept of a singular point has already appeared in various parts of our course. For example, a detailed discussion of isolated singular points of single-valued character can be found in Chap. II.1, while algebraic and logarithmic branch points are encountered in Chap. I.11. In each of these cases, the analyticity of the given function is violated in a particularly simple way. Thus the function fails to be continuous at a pole or even wide-sense continuous at an essential singular point, while the function fails to be single-valued in a neighborhood of a branch point. The concept of a singular point of a circular element $e_0 = \{K, f(z)\}$ (introduced in Sec. I.84) is of a more general nature, i.e., a point ζ on the boundary of the disk K is said to be a singular point of e_0 if there is no single-valued analytic function $\varphi_\zeta(z)$ defined in a neighborhood $\mathcal{N}(\zeta)$ such that $\varphi_\zeta(z) = f(z)$ for all $z \in \mathcal{N}(\zeta) \cap K$. In special cases, a singular point of this kind may be a pole, an essential singular point or a branch point (recall the examples given in Sec. I.84).

The time has now come to use analytic continuation to develop a unified theory of singular points, comprising all the special kinds of singular points just described:

DEFINITION 1. *Let e_0 be a circular element with center z_0, and let L be a continuous curve with equation*

$$z = \lambda(t) \qquad (\alpha \leqslant t \leqslant \beta)$$

and initial point $z_0 = \lambda(\alpha)$. If e_0 can be continued along the whole curve L, then every point of L is said to be a regular point (relative to the given continuation). Otherwise, there exists a number $\tau \in (\alpha, \beta]$ such that e_0 can be continued along every subarc $L[\alpha, t]$, $t < \tau$ but not along $L[\alpha, \tau]$ itself. In this case, the point $\zeta = \lambda(\tau) \in L[\alpha, \tau]$ is called a singular point (again relative to the given continuation[28]), but every other point of $L[\alpha, \tau]$ is called regular.

Thus, figuratively speaking, a singular point of an analytic function is an obstacle to analytic continuation along some path of the elements defining the function. The reader can easily verify that poles, essential singular points and limit points of poles are all singular points in the sense of the new definition, and that so are all singular points of circular elements as defined in Sec. I. 84 (do this in detail, thinking about the role of the particular element and path chosen). The case of branch points (discussed so far only

[28] The full significance of this stipulation will emerge later (see Definition 2, p. 283).

for particularly simple examples like $z^{1/n}$, Ln z, etc.) warrants further study, which we carry out in a series of theorems.

THEOREM 8.8. *Let G be an annular domain of the form $0 < |z - \zeta| < \rho$ (a deleted neighborhood) containing a circular element e_0 with center z_0. Suppose e_0 can be continued along every arc $\overset{\frown}{z_0 z} \subset G$, thereby defining a function $f(z)$ at every point of G, where $f(z)$ is in general multiple-valued. Then the function*

$$f^*(t) = f(\zeta + e^t) \tag{8.17}$$

is single-valued and analytic (but in general nonperiodic!) in the half-plane D: Re $t <$ ln ρ.

Proof. Every single-valued branch of the multiple-valued function

$$t = \text{Ln}\,(z - \zeta) \tag{8.18}$$

carries G into a horizontal strip of width 2π in the t-plane, and the set of all such branches carries G into D. Let $e_0 = \{K, f_0\}$ and choose a fixed branch of (8.18), mapping z_0 into t_0. Then $\varepsilon_0 = \{\varkappa, f_0(\zeta + e^t)\}$ is itself a circular element if its disk $\varkappa$, centered at t_0, is chosen sufficiently small. Since e_0 can be continued along every curve $\overset{\frown}{z_0 z} \subset G$, the element ε_0 can be continued along every curve $\overset{\frown}{t_0 t} \subset D$ (why?), thereby defining a function $f^*(t)$ on D. According to the monodromy theorem and its corollary, $f^*(t)$ is single-valued and analytic on D. Moreover, given any point $t' \in D$, let $z' \in G$ be the inverse image of t' under the mapping $t = \text{Ln}\,(z - \zeta)$. Then $f(z)$ clearly has a branch $\tilde{f}(z)$ such that $f^*(t') = \tilde{f}(z')$, and this is precisely what is meant by (8.17).

THEOREM 8.9. *With the same notation as in Theorem 8.8, suppose $f^*(t)$ has a period of the form $2k\pi i$, where k is a positive integer, but no period of the form $2j\pi i$, where j is a positive integer smaller than k. Then the result of continuing any element of $f(z)$, in particular e_0, along any closed Jordan curve γ with $\zeta \in I(\gamma)$ is a new element with the same center, and the result of k consecutive continuations (but no fewer) along γ in the same direction is the original element. On the other hand, the result of continuing an element of $f(z)$ along any closed Jordan curve γ with $\zeta \notin I(\gamma)$ is the original element.*

Proof. Clearly

$$f^*(t + 2k\pi i) = f^*(t)$$

for all $t \in D$, while

$$f^*(t + 2j\pi i) = f^*(t) \qquad (0 < j < k) \tag{8.19}$$

cannot hold even for infinitely many values of t with a limit point in D, since otherwise the interior uniqueness theorem would imply (8.19) for all t, and then $2j\pi i$ would be a period of $f^*(t)$ smaller than $2k\pi i$, contrary

to hypothesis. The transition from $f^*(t_0)$ to $f^*(t_0 + 2k\pi i)$, where t_0 is any image of z_0 under the mapping $t = \operatorname{Ln}(z - \zeta)$, can be achieved by analytic continuation of an element ε_0 of $f^*(t)$ with center t_0 along any curve in D joining t_0 to $t_0 + 2k\pi i$, in particular along the line segment L joining t_0 to $t_0 + 2k\pi i$. Let γ be the inverse image under the mapping $t = \operatorname{Ln}(z - \zeta)$ of the segment joining t_0 to $t_0 + 2\pi i$. Then obviously $z_0 \in \gamma$, $\zeta \in I(\gamma)$,[29] and as t traverses L, the quantity

$$\operatorname{Arg}(z - \zeta) = \operatorname{Im} \operatorname{Ln}(z - \zeta) = \operatorname{Im} t$$

increases by $2k\pi$, corresponding to k circuits of the point z around γ in the positive (counterclockwise) direction. By the same token, continuing ε_0 along L corresponds to k consecutive continuations of an element e_0 with center z_0 along γ in the positive direction, resulting in an element e_0' with center z_0. But, by the periodicity of $f^*(t)$, the values of e_0' must be the same as those of e_0 in a neighborhood of z_0, and hence the elements e_0 and e_0' coincide (in the sense that one is subordinate to the other). Moreover, γ can be replaced by an arbitrary closed Jordan curve by suitably modifying the curve joining t_0 to $t_0 + 2k\pi i$. The same argument applies equally well to any element e_1 with center $z_1 \neq z_0$, obtained by continuing e_0 along a curve $\overset{\frown}{z_0 z_1} \subset G$, i.e., the result of k consecutive continuations of e_1 along a closed Jordan curve γ_1 with $\zeta \in I(\gamma_1)$ is again the element e_1.

Clearly, there cannot exist an element $\tilde{e}$ and a corresponding closed Jordan curve $\tilde{\gamma}$ with $\zeta \in I(\tilde{\gamma})$ such that the result of fewer than k continuations along $\tilde{\gamma}$ in the positive direction is again the element $\tilde{e}$. In fact, (8.19) would then hold for infinitely many values of t with a limit point (the center of $\tilde{e}$) in D, which is impossible for the reason already given. To verify the last assertion of the theorem, we need only note that the image under the mapping $t = \operatorname{Ln}(z - \zeta)$ of a closed Jordan curve γ with $\zeta \notin I(\gamma)$ is a closed Jordan curve $\Gamma \subset D$ (why?). But continuation along Γ cannot alter an element of the function $f^*(t)$, and hence continuation along γ cannot alter an element of the function $f(z)$. Finally, we observe that γ can be traversed in the negative direction instead of the positive direction.[30]

COROLLARY 1. *If $k = 1$, the function $f(z)$ is single-valued and analytic on G. Then ζ is a regular point of $f(z)$ if e_0 can be continued along every curve $\overset{\frown}{z_0\zeta}$ contained in G (except for the final point ζ), while ζ is an isolated singular point of single-valued character if e_0 cannot be continued along*

[29] In fact, γ is a circle with center ζ.

[30] More generally, γ need not be a Jordan curve, different "loops" of γ need not coincide, and there may be winding around ζ in both directions provided k "uncompensated" circuits are made in one direction (examine the various possibilities).

some such curve.[31] *In particular, $f(z)$ has a Taylor or Laurent expansion at ζ.*

Proof. If $k = 1$, $f^*(t)$ has period $2\pi i$. Then continuation along an arbitrary closed Jordan curve $\gamma \subset G$ cannot alter any element of $f(z)$.

COROLLARY 2. *If $k > 1$, the function $f(z)$ is k-valued and analytic on G, in the sense that it has precisely k distinct elements at every point $z \in G$. Moreover, ζ is a branch point of order $k - 1$ of $f(z)$.*

Proof. If $k > 1$, $f^*(t)$ has period $2k\pi i$. Given any point $t \in D$, consider the sequences of coefficients in the Taylor expansions of $f^*(t)$ at the points

$$t, \quad t + 2\pi i, \ldots, \quad t + 2(k-1)\pi i. \tag{8.20}$$

No two sequences can coincide completely, since otherwise $f^*(t)$ would have a smaller period $2j\pi i$ $(0 < j < k)$. Therefore the k elements of $f^*(t)$ centered at the points (8.20) are distinct, and hence so are the corresponding k elements of $f(z)$, all centered at the same point $z = \zeta + e^t$. Moreover, since $2k\pi i$ is a period of $f^*(t)$, there can be no more than k distinct elements at z. Let $e^{(0)}$ be any of these elements, let γ be a closed Jordan curve such that $z \in \gamma$, $\zeta \in I(\gamma)$, and let $e^{(n)}$ be the result of $n > 0$ consecutive continuations of $e^{(0)}$ along γ in the same direction. Then

$$e^{(0)}, \quad e^{(1)}, \ldots, \quad e^{(k-1)} \tag{8.21}$$

are the k distinct elements of $f(z)$ at z, and moreover

$$e^{(n+k)} = e^{(n)} \qquad (n = 0, 1, \ldots, k-1).$$

By the monodromy theorem and its corollary, each element (8.21) defines a distinct single-valued analytic branch of $f(z)$ on the simply connected domain G' obtained from G by deleting any Jordan curve in the extended z-plane joining the points 0 and ∞. The last assertion of the corollary now follows at once from the considerations on p. I.217.

COROLLARY 3. *At a branch point ζ of order $k - 1$, $f(z)$ has an expansion of the form*

$$f(z) = \sum_{n=-\infty}^{\infty} a_n (z - \zeta)^{n/k} \qquad (z \in G). \tag{8.22}$$

Proof. The substitution

$$Z = (z - \zeta)^{1/k} \tag{8.23}$$

transforms $G: 0 < |z - \zeta| < \rho$ into the annular domain $G_1: 0 < |Z| < \rho^{1/k}$ and $f(z)$ into the function

$$f_1(Z) = f(\zeta + Z^k),$$

[31] See Problem 8.11.

defined by continuation of any of its elements along all curves contained in G_1. Then the substitution

$$\tau = \operatorname{Ln} Z = \frac{1}{k} \operatorname{Ln} (z - \zeta) = \frac{1}{k} t$$

transforms G_1 into the half-plane D_1: $\operatorname{Re} \tau < (1/k) \ln \rho$ and $f_1(Z)$ into the single-valued analytic function

$$f_1^*(\tau) = f^*(k\tau)$$

with period $2\pi i$. Therefore, by Corollary 1, $f_1(Z)$ is single-valued and analytic on G_1, with a Laurent expansion of the form

$$f_1(Z) = \sum_{n=-\infty}^{\infty} a_n Z^n \qquad (Z \in G_1), \tag{8.24}$$

and (8.22) is an immediate consequence of (8.23) and (8.24).

Let ζ be a branch point of order $k - 1$ of $f(z)$, and suppose $f(z)$ has a finite or infinite limit at $z = \zeta$. Then $f_1(Z)$ has a finite or infinite limit at $Z = 0$, and hence all but a finite number of coefficients a_n with negative indices vanish. In this case, ζ is called an *algebraic branch point*, just as on p. I.217. However, suppose $f(z)$ has no limit at $z = \zeta$. Then $f_1(Z)$ has no limit at $Z = 0$, and hence there are infinitely many nonzero coefficients a_n with negative indices. In this case, ζ is called a *transcendental branch point* (of finite order).[32]

Example 1. Each of the functions $z^{1/k}$ and $\exp(z^{1/k})$ has an algebraic branch point of order $k - 1$ at the origin.

Example 2. The function $\exp(z^{-1/k})$ has a transcendental branch point of order $k - 1$ at the origin.

We must still discuss the case where $f^*(t)$ has no periods which are integral multiples of $2\pi i$:

THEOREM 8.10. *With the same notation as in Theorem* 8.8, *suppose* $f^*(t)$ *has no periods of the form* $2k\pi i$, *where* k *is a positive integer. Then the result of continuing an element of* $f(z)$, *in particular* e_0, *along any closed Jordan curve* γ *with* $\zeta \in I(\gamma)$ *is a new element with the same center, and the result of any number of consecutive continuations along* γ *in the same direction is always a new element. On the other hand, the result of continuing an element of* $f(z)$ *along any closed Jordan curve* γ *with* $\zeta \notin I(\gamma)$ *is the original element.*

Proof. Use the second paragraph of the proof of Theorem 8.9, replacing k by ∞ and the condition $0 < j < k$ figuring in (8.19) by $j > 0$.

---

[32] Logarithmic branch points are also classified as transcendental branch points (of infinite order).

COROLLARY 1. *The function $f(z)$ is infinite-valued and analytic on G, in the sense that it has infinitely many distinct elements at every point $z \in G$. Moreover, ζ is a logarithmic branch point of $f(z)$.*

Proof. Instead of (8.20), consider the sequences of coefficients in the Taylor expansions of $f^*(t)$ at the points

$$t, \quad t \pm 2\pi i, \quad t \pm 4\pi i, \ldots \tag{8.25}$$

No two sequences can coincide completely, since otherwise $f^*(t)$ would have a period of the form $2j\pi i$. Therefore the infinitely many elements of $f^*(t)$ centered at the points (8.24) are distinct, and hence so are the corresponding infinitely many elements of $f(z)$, all centered at the same point $z = \zeta + e^t$. Let $e^{(0)}$ be any of these elements, and let γ be a closed Jordan curve such that $z \in \gamma$, $\zeta \in I(\gamma)$. Moreover, let $e^{(n)}$ be the result of $n > 0$ consecutive continuations of $e^{(0)}$ along γ in the positive direction, and let $e^{(-n)}$ be the result of n consecutive continuations in the negative direction. Then

$$\ldots, e^{(-n)}, \ldots, e^{(-1)}, e^{(0)}, e^{(1)}, \ldots, e^{(n)}, \ldots \tag{8.26}$$

are the distinct elements of $f(z)$ at z, and any of these elements can be obtained from any other element by making a suitable number of continuations along γ in the same direction (whether positive or negative). By the monodromy theorem and its corollary, each element (8.26) defines a distinct single-valued analytic branch of $f(z)$ on the simply connected domain G' obtained from G by deleting any Jordan curve in the extended z-plane joining the points 0 and ∞. The last assertion of the corollary now follows at once from the considerations on pp. I.217 and I.228.

COROLLARY 2. *At a logarithmic branch point ζ, $f(z)$ has an expansion of the form*

$$f(z) = \sum_{n=0}^{\infty} a_n \left[\frac{\operatorname{Ln}(z - \zeta) - \alpha}{\operatorname{Ln}(z - \zeta) + \alpha}\right]^n, \tag{8.27}$$

where α is any point of D.

Proof. The substitution

$$\tau = \frac{t - \alpha}{t + \alpha} \qquad (\alpha \in D) \tag{8.28}$$

transforms the half-plane D into the unit disk $|\tau| < 1$ and $f^*(t)$ into the function

$$F(\tau) = f^*\left(\alpha \frac{1 + \tau}{1 - \tau}\right),$$

which is single-valued and analytic on K. Therefore $F(\tau)$ has a Taylor expansion of the form

$$F(\tau) = \sum_{n=0}^{\infty} a_n \tau^n \qquad (\tau \in K),$$

which together with (8.28) implies first

$$f^*(t) = \sum_{n=0}^{\infty} a_n \left(\frac{t-\alpha}{t+\alpha}\right)^n$$

and then (8.27).

The expansion (8.27) is not particularly convenient for practical purposes. However, much simpler expansions of a function $f(z)$ in a neighborhood of a logarithmic branch point ζ can often be found:

Example 1. Suppose one circuit around ζ [more precisely, around a closed Jordan curve γ with $\zeta \in I(\gamma)$] in the positive direction causes $f(z)$ to increase by a constant A. Then the circuit around ζ has no effect on the function

$$\varphi(z) = f(z) - \frac{A}{2\pi i} \operatorname{Ln}(z - \zeta),$$

i.e., $\varphi(z)$ is single-valued and analytic in a deleted neighborhood of ζ. Therefore $f(z)$ has an expansion of the form

$$f(z) = \frac{A}{2\pi i} \operatorname{Ln}(z - \zeta) + \sum_{n=-\infty}^{\infty} a_n (z - \zeta)^n$$

at ζ.

Example 2. Suppose one circuit around ζ causes $f(z)$ to be multiplied by a constant $a \neq 0$. Then the circuit around ζ has no effect on the function

$$\varphi(z) = f(z) \exp\left\{-\frac{\ln a}{2\pi i} \operatorname{Ln}(z - \zeta)\right\} = f(z)(z - \zeta)^{-\ln a/2\pi i},$$

and hence $f(z)$ has an expansion of the form

$$f(z) = (z - \zeta)^{\ln a/2\pi i} \sum_{n=-\infty}^{\infty} a_n (z - \zeta)^n$$

at ζ.

Remark. So far it has been tacitly assumed that the branch point ζ is finite. However, the case $\zeta = \infty$ can be treated by the usual device, i.e., by making the preliminary transformation $Z = 1/z$ and then considering the function

$$f^*(Z) = f\left(\frac{1}{Z}\right).$$

In this sense, all the results obtained above remain valid if we replace $z - \zeta$ by $1/Z$.

Definition 1, p. 276 is still incomplete, since it does not say when two singular points defined by different continuations are regarded as identical. The appropriate compatibility conditions are given by

DEFINITION 2. *Let e_1 and e_2 be two circular elements of a complete analytic function $F(z)$, with centers z_1 and z_2. Let*

$$L_1: z = \lambda_1(t_1), \qquad L_2: z = \lambda_2(t_2)$$

be two continuous curves, with initial points z_1 and z_2, which have at least one point ζ in common, corresponding to the parameter values $t_1 = \tau_1$, $t_2 = \tau_2$, where ζ is a singular point relative to both continuation of e_1 along L_1 and continuation of e_2 along L_2. Given any neighborhood $\mathcal{N}(\zeta)$, suppose there exists a number $\varepsilon > 0$ such that the result of continuing e_1 along L_1 up to any point

$$z' = \lambda_1(t'), \qquad \tau_1 - \varepsilon < t' < \tau_1 \tag{8.29}$$

and then along any continuous arc $\overset{\frown}{z'z''} \subset \mathcal{N}(\zeta)$, where

$$z'' = \lambda_2(t''), \qquad \tau_2 - \varepsilon < t'' < \tau_2, \tag{8.30}$$

is the same as the result of continuing e_2 along L_2 up to the point z'', and vice versa.[33] *Then $\lambda_1(\tau_1)$ and $\lambda_2(\tau_2)$ are regarded as identical, i.e., as corresponding to a unique singular point ζ of $F(z)$. Otherwise, $\lambda_1(\tau_1)$ and $\lambda_2(\tau_2)$ are regarded as distinct singular points of $F(z)$, represented by the same point ζ of the complex plane.*

Example. Consider the function

$$F(z) = \frac{1}{(1 + \sqrt{z})(1 + \sqrt[6]{z})},$$

and let $L_1 = L_2$ be the unit circle traversed once in the positive direction starting from the point $z = z_1 = z_2 = 1$. Let e_1 denote an element with its center at $z = 1$, representing the single-valued branch of $F(z)$ specified in a neighborhood of $z = 1$ by the conditions

$$\sqrt{z} = \sqrt{|z|}\left(\cos\frac{\arg z}{2} + i\sin\frac{\arg z}{2}\right),$$

$$\sqrt[6]{z} = \sqrt[6]{|z|}\left(\cos\frac{\arg z}{6} + i\sin\frac{\arg z}{6}\right).$$

Then, as a result of continuation along L_1, $\sqrt{z}$ goes into [34]

$$-\sqrt{|z|}\left(\cos\frac{\arg z}{2} + i\sin\frac{\arg z}{2}\right), \tag{8.31}$$

---

[33] Explicitly, the result of continuing e_2 along L_2 up to any point (8.30) and then along any arc $\overset{\frown}{z''z'} \subset \mathcal{N}(\zeta)$, where z' is given by (8.29), is the same as the result of continuing e_1 along L_1 up to the point z'.

[34] As a result of the analytic continuation, each branch of $\sqrt{z}$ (and of $\sqrt[6]{z}$) changes continuously into the "next branch," as in Sec. I.54.

and hence $F(z)$ becomes infinite as $z \to 1$ along L_1. It follows that $z = 1$ is a singular point relative to this continuation.

Next let e_2 denote an element with its center at $z = 1$, representing the single-valued branch of $F(z)$ specified in a neighborhood of $z = 1$ by the conditions

$$\sqrt{z} = \sqrt{|z|}\left(\cos\frac{\arg z}{2} + i\sin\frac{\arg z}{2}\right),$$

$$\sqrt[6]{z} = \sqrt[6]{|z|}\left(\cos\frac{\arg z + 4\pi}{6} + i\sin\frac{\arg z + 4\pi}{6}\right).$$

Then, as a result of continuation along L_2, $\sqrt{z}$ goes into (8.31) again, while $\sqrt[6]{z}$ goes into

$$-\sqrt[6]{|z|}\left(\cos\frac{\arg z}{6} + i\sin\frac{\arg z}{6}\right).$$

Therefore $F(z)$ becomes infinite as $z \to 1$ along L_2, and hence $z = 1$ is a singular point relative to this continuation as well. However, it is easy to see that the two singular points of $F(z)$ obtained in this way are distinct, although represented by the same point $z = 1$. In fact, the values of $F(z)$ resulting from continuation of the two elements e_1 and e_2 along $L_1 = L_2$ belong to two distinct single-valued branches of $F(z)$ in a neighborhood of $z = 1$, i.e.,

$$F_1(z) = \frac{1}{1 - \sqrt{|z|}\left(\cos\frac{\arg z}{2} + i\sin\frac{\arg z}{2}\right)} \times \frac{1}{1 + \sqrt[6]{|z|}\left(\cos\frac{\arg z + 2\pi}{6} + i\sin\frac{\arg z + 2\pi}{6}\right)}$$

and

$$F_2(z) = \frac{1}{1 - \sqrt{|z|}\left(\cos\frac{\arg z}{2} + i\sin\frac{\arg z}{2}\right)} \times \frac{1}{1 - \sqrt[6]{|z|}\left(\cos\frac{\arg z}{6} + i\sin\frac{\arg z}{6}\right)}.$$

But it is impossible to go from one of these branches to the other by analytic continuation along a curve contained in a sufficiently small neighborhood of $z = 1$ (here "sufficiently small" means of radius no greater than 1). Therefore $z = 1$ represents two distinct singular points of $F(z)$, as asserted. Clearly, the first of these singular points is a pole of order 1, while the second is a pole of order 2.

The expansion (8.22) valid at an algebraic singular point suggests that we extend the concept of a circular element by considering (*generalized*) *elements* of the form

$$e: \sum_{n=m}^{\infty} a_n(z - z_0)^{n/k} \qquad (|z - z_0| < r) \tag{8.32}$$

or

$$e: \sum_{n=m}^{\infty} a_n z^{-n/k} \qquad (|z| > r), \tag{8.33}$$

where m is an arbitrary integer and k a positive integer (so far we have only dealt with the case $m \geqslant 0$, $k = 1$). As in Definition 1, p. 264, an element of the form (8.32) is said to have series

$$\sum_{n=m}^{\infty} a_n(z - z_0)^{n/k},$$

center z_0, radius r and disk $|z - z_0| < r$, while an element of the form (8.33) is said to have series

$$\sum_{n=m}^{\infty} a_n z^{-n/k},$$

center ∞, radius r and "disk" $|z| > r$.

The generalized elements (8.32) and (8.33) are said to be in *canonical form* if $a_m \neq 0$ and if the fractions n/k corresponding to nonzero a_n are not all of the form $n'q/k'q$, where q is an integer greater than 1. An element written in canonical form is said to be *unbranched* if $k = 1$ and *branched* (*of order* $k - 1$) if $k > 1$. By a *regular* element we mean an unbranched element with $m \geqslant 0$, and any other kind of generalized element, branched or not, will be called *critical.*

The logical next step is to extend the concept of analytic continuation to the case of generalized elements. Here it is clearly sufficient to consider the case of branched elements. To be explicit, let e be the element (8.32) written in canonical form with $k > 1$, let $K_0: |z - z_0| < r$ be the disk of e, and let $F(z)$ be the sum of the series of e. Given any $z_1 \in K_0$ ($z_1 \neq z_0$), let

$$\rho = \min\{|z_1 - z_0|, r - |z_1 - z_0|\}.$$

Then $F(z)$ has k single-valued branches on the disk $K_1: |z - z_1| < \rho$. To fix one of these branches, we choose one of the k single-valued branches of the function $(z - z_0)^{1/k}$. Every element of the form $\{K_1, f(z)\}$, where $f(z)$ is one of the k single-valued branches of $F(z)$, is said to be *subordinate* to the original element e. Clearly, every subordinate element of the branched element e is itself unbranched, and in fact regular, since $F(z)$ is finite except at the center of e.

DEFINITION 3. *Two generalized elements of the form* (8.32) *or* (8.33) *are said to be direct* (*analytic*) *continuations of each other if they have at*

least one subordinate element in common. They are said to be (analytic) continuations of each other if they can be joined by a chain of direct continuations, just as in Definition 3, *p.* 258.

Next we introduce the analogue of a complete analytic function, for the case where critical elements are permitted:

DEFINITION 4. *Let $\mathscr{K}$ be a class of generalized elements such that two elements belong to $\mathscr{K}$ if and only if they are continuations of each other. Then $\mathscr{K}$ is said to define an analytic configuration, and distinct classes define distinct analytic configurations. Every generalized element in $\mathscr{K}$ is called an element of the analytic configuration.*

Remark. Suppose all critical elements are removed from $\mathscr{K}$. Then the remaining class of elements defines a complete analytic function in the sense of Definition 4, p. 259. Conversely, given a complete analytic function $F(z)$, we obtain an analytic configuration by adjoining to $F(z)$ all critical elements which are continuations of the elements of $F(z)$.

Example 1. The class of all power series representing single-valued branches of the function $\sqrt[k]{z}$, i.e., the class of all series of the form

$$\sqrt[k]{z_0} \sum_{n=0}^{\infty} \binom{1/k}{n} \left(\frac{z - z_0}{z_0}\right)^n,$$

where z_0 is an arbitrary complex number different from 0 or ∞, defines a complete analytic function. To obtain the corresponding analytic configuration, representing the multiple-valued function $\sqrt[k]{z}$ on its whole domain of definition (the extended complex plane), we need only adjoin two branched elements, both represented by the same degenerate series $\sqrt[k]{z}$, but one corresponding to a neighborhood of $z = 0$ and the other to a neighborhood of $z = \infty$.

Example 2. The class of all power series representing single-valued branches of the function Arc sin z defines a complete analytic function.[35] By adjoining the infinite set of branched elements representing Arc sin z at the points ± 1 (see Problem 8.20), we obtain the analytic configuration representing the multiple-valued function Arc sin z on its whole domain of definition (the finite z-plane).

Remark. In the case of functions meromorphic in the whole plane, the analytic configuration coincides with the complete analytic function. The same is true for any multiple-valued function defined as the inverse of a meromorphic function, if its Riemann surface has no branch points.

---

[35] The branch points of Arc sin z lie over the points ± 1, as in Example 3, pp. 251–252.

43. The Analytic Configuration as a Topological Surface

Let A be an analytic configuration, as in Definition 4 above. Suppose two elements of A are regarded as identical if they have the same center and the same series, regardless of the size of their disks (which are concentric with the circle of convergence of the series). Then A is easily made into a topological space, with the elements of A as the points of the space and the "natural" definition of neighborhoods:

DEFINITION. *Let A be an analytic configuration, and let the element* (8.32) [*or* (8.33)] *be a point of A. Suppose $|z - z_0| = R$ (or $|z| = R$) is the circle of convergence of the element, and let $\rho \leqslant R$ (or $\rho \geqslant R$ if $z_0 = \infty$). Then by a ρ-neighborhood of e is meant the set of all (generalized) elements subordinate to e (including e itself) whose centers belong to the disk $|z - z_0| < \rho$ (or $|z| > \rho$ if $z_0 = \infty$).*

This definition immediately implies

THEOREM 8.11. *The analytic configuration A is a Hausdorff space.*

Proof. Recall Definitions 1 and 2, p. 217.

Actually, it turns out that A is a topological surface, as we now proceed to show in gradual stages:

THEOREM 8.12. *The analytic configuration A is locally Euclidean.*

Proof. First let e_0 be an unbranched element, with disk K and center z_0. Then there is a unique subordinate element $e \subset e_0$ with its center at any given point $z \in K$. Given any neighborhood $\mathcal{N}(e_0; \rho)$ and any $e \in \mathcal{N}(e_0; \rho)$,[36] let $z = \varphi_{e_0}(e)$ be the mapping which assigns e its center z. Then $\varphi_{e_0}(e)$ maps $\mathcal{N}(e_0; \rho)$ onto the disk $|z - z_0| < \rho$ (here we assume for definiteness that z_0 is finite), and the mapping is obviously one-to-one. Moreover $\varphi_{e_0}(e)$ is continuous in both directions, since

$$e_1 \in \mathcal{N}(e_0; \rho), \quad z_1 = \varphi_{e_0}(e_1), \quad e \in \mathcal{N}(e_1; \rho_1), \quad z = \varphi_{e_1}(e),$$

where

$$\rho_1 = \min \{|z_1 - z_0|, \rho - |z_1 - z_0|\}$$

implies that the disk $|z - z_1| < \rho_1$ is contained in the disk $|z - z_0| < \rho$, and conversely. Therefore $z = \varphi_{e_0}(e)$ is a homeomorphic mapping of $\mathcal{N}(e_0; \rho)$ onto the disk $|z - z_0| < \rho$.

---

[36] $\mathcal{N}(e_0; \rho)$ is the set of all elements of A subordinate to e_0 with centers at points of the disk $|z - z_0| < \rho$ if $z_0 \neq \infty$, and at points of the "disk" $|z| > \rho$ if $z_0 = \infty$. If ρ is unimportant, we simply write $\mathcal{N}(e_0)$.

Now let

$$e_0: \sum_{n=m}^{\infty} a_n(z - z_0)^{n/k} \qquad (|z - z_0| < \rho) \tag{8.34}$$

be a branched element of order $k - 1$, written in canonical form (to be explicit, we again assume that z_0 is finite). Given any point $z_1 \neq z_0$ in the disk $|z - z_0| < \rho$, there are now several (in fact $k > 1$) distinct elements subordinate to e_0, each with its center at z_1. Therefore the function mapping e into its center is no longer one-to-one on $\mathcal{N}(e_0)$. To find a homeomorphic mapping of $\mathcal{N}(e_0)$ onto a disk, we introduce the parameter

$$t = (z - z_0)^{1/k}. \tag{8.35}$$

If $z_1 \neq z_0$, the function (8.35) takes k distinct values $t_1, \ldots, t_k$ at the point z_1, each of which is carried into any other by a rotation around the point z_0 through an angle equal to a suitable integral multiple of $2\pi/k$. Correspondingly, these rotations carry any sufficiently small neighborhood of z_1 (which, in particular, does not contain z_0) into k disjoint domains $g_1, \ldots, g_k$, where g_j is uniquely characterized by the condition that it contain the point t_j. Clearly, each of the domains $g_1, \ldots, g_k$ is carried into any other by a rotation around z_0 through an angle equal to a suitable integral multiple of $2\pi/k$. To specify each of the k single-valued analytic branches of (8.35) defined in a neighborhood of z_1, we need only give the domain g_j in which the branch takes (some of) its values, or equivalently the value t_j which the branch takes at $z = z_1$. In just the same way, we specify each of the k single-valued analytic branches of the k-valued sum of the series in (8.34), and hence one of the k elements $e_1^{(j)}$ $(j = 1, \ldots, k)$ subordinate to e_0, each with its center at z_1.

Next we observe that two elements $e_1^{(j)}$ and $e_1^{(j+p)}$ corresponding to distinct values t_j and t_{j+p} $(p < k)$ are themselves distinct. In fact, suppose this is not true, i.e., suppose $e_1^{(j)}$ and $e_1^{(j+p)}$ coincide. In a neighborhood of z_1, $e_1^{(j)}$ and $e_1^{(j+p)}$ are represented by the series

$$\sum_{n=m}^{\infty} a_n(z - z_0)^{n/k}, \tag{8.36}$$

where $t = (z - z_0)^{1/k}$ range over the domains g_j and g_{j+p}, respectively. Therefore, in terms of the variable t, $e_1^{(j)}$ and $e_1^{(j+p)}$ are represented by the series

$$\sum_{n=m}^{\infty} a_n t^n,$$

where $t \in g_j$ and $t \in g_{j+p}$. Given any z in a neighborhood of z_1, let $t = (z - z_0)^{1/k}$ be the corresponding point of g_j. Then the number

$e^{2\pi ip/k}t \in g_{j+p}$ also corresponds to z_j, and hence the series representing $e_1^{(j)}$ and $e_1^{(j+p)}$ *on the domain* g_j are

$$\sum_{n=m}^{\infty} a_n t^n \quad \text{and} \quad \sum_{n=m}^{\infty} a_n e^{2\pi ipn/k} t^n.$$

Since $e_1^{(j)}$ and $e_1^{(j+p)}$ coincide, it follows from Theorem I.16.6 that

$$e^{2\pi ipn/k} = 1$$

for all n such that $a_n \neq 0$. But then the pn/k are all integers, and hence the fractions n/k corresponding to nonzero a_n are all of the form $n'q/k'q$, where q is an integer greater than 1 (here we use the fact that $p < k$). This contradicts the assumption that (8.34) is written in canonical form, thereby proving that the elements $e_1^{(j)}$ and $e_1^{(j+p)}$ are indeed distinct.

Now let $\varphi_{e_0}(e)$ be the function assigning the value t_j to $e_1^{(j)}$ ($j = 1, \ldots, p$) and similarly for every $e \in \mathcal{N}(e_0; \rho)$, while assigning the value 0 to the branched element e_0 itself. Then $\varphi_{e_0}(e)$ maps $\mathcal{N}(e_0; \rho)$ onto the disk $|t| < \rho^{1/k}$, and the mapping is obviously one-to-one. Moreover, $\varphi_{e_0}(e)$ is continuous in both directions. In fact, the correspondence between t and z established by every single-valued branch of (8.35) in a neighborhood of a point $z_1 \neq z_0$ is continuous in both directions, and so is the correspondence between the elements subordinate to e_0 representing the various single-valued branches of the sum of the series (8.36) in a neighborhood of z_1 and their centers (belonging to a neighborhood of z_1 which does not contain z_0). The continuity remains in force both at $e = e_0$ and at $t = 0$. In fact, as $t \to 0$, the corresponding elements approach e_0, i.e., lie in any preassigned neighborhood $\mathcal{N}(e_0)$ for all sufficiently small $|t|$, while conversely, as $e \to e_0$, the centers of the elements e approach z_0 and hence $t \to 0$.

Thus, finally, if e_0 is branched and if e is an element in $\mathcal{N}(e_0)$ with center z, the function

$$t = \varphi_{e_0}(e) = (z - z_0)^{1/k} \tag{8.37}$$

maps $\mathcal{N}(e_0)$ homeomorphically onto a disk in the t-plane centered at the origin. Associating a mapping of the form (8.37) with every branched element e_0, and a mapping of the form $z = \varphi_{e_0}(e)$ with every unbranched e_0, we obtain a set of local parameters like those figuring in Definition 3, p. 218. Therefore A is locally homeomorphic to the plane, and the proof is complete.

LEMMA 1. *Given any generalized element* e_0 *with disk* K, *let* ε *and* ε' *be any two elements subordinate to* e_0 (*and hence regular*). *Then* ε' *is the continuation of* ε *along a curve* $L \subset K$, *where the chain of elements joining* ε *to* ε' *along* L *consists entirely of regular elements subordinate to* e_0.

Proof. If e_0 is regular, the proof is an immediate consequence of the lemma on p. 267, since any two points of K can be joined by a (Jordan) curve contained in K (recall Theorem I.4.12). Thus suppose e_0 is critical, in particular branched, with series

$$\sum_{n=m}^{\infty} a_n(z - z_0)^{n/k}, \tag{8.38}$$

written in canonical form (clearly this case is sufficiently general for the purposes of our proof). The transformation $t = (z - z_0)^{1/k}$ carries e_0 into an unbranched element $\tilde{e}_0$ with disk $\tilde{K}$ centered at $t = 0$, while at the same time ζ and ζ', the centers of ε and ε', go into points τ and τ' in $\tilde{K}$. Let $\tilde{L} \subset \tilde{K}$ be a continuous curve joining τ to τ' without passing through the center of $\tilde{K}$ (the point $t = 0$), and let $\tilde{\varepsilon}_0, \tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$ be a chain of elements with centers $\tau = \tau_0, \tau_1, \ldots, \tau_n = \tau'$, all subordinate to $\tilde{e}_0$, joining $\tilde{\varepsilon}_0$ to $\tilde{\varepsilon}_n$ along $\tilde{L}$ (the existence of such a chain follows from the lemma on p. 267. Under the mapping

$$z = z_0 + t^k, \tag{8.39}$$

the curve $\tilde{L}$ goes into a curve $L \subset K$ in the z-plane which joins ζ to ζ' without passing through z_0, and the disks of the elements $\tilde{\varepsilon}_j$ go into simply connected domains g_j $(j = 1, \ldots, n)$ where each $g_{j-1} \cap g_j$ is a nonempty (connected) set covering the arc $\sigma_j \subset L$ connecting the image of τ_{j-1} to the image of τ_j. The transformation (8.39) carries the elements $\tilde{\varepsilon}_j$ themselves into (noncircular) elements $\{g_j, \varphi_j(z)\}$, all subordinate to e_0, where each $\varphi_j(z)$ is a single-valued branch of the sum of the series (8.38). Clearly, $\{g_0, \varphi_0\}$ is a direct continuation of ε and $\{g_n, \varphi_n\}$ is a direct continuation of ε', while each $\{g_j, \varphi_j\}$ $(j = 1, \ldots, n)$ is a direct continuation of $\{g_{j-1}, \varphi_{j-1}\}$.

Now choose a point $\zeta_j \in L$ in each domain $g_{j-1} \cap g_j$, and set $\zeta_0 = \zeta$, $\zeta_{n+1} = \zeta'$. Let ε_j $(j = 1, \ldots, n)$ be the element with center ζ_j which is subordinate to both $\{g_{j-1}, \varphi_{j-1}\}$ and $\{g_j, \varphi_j\}$ (and hence to e_0). Moreover, let ε_0 be the element with center ζ_0 subordinate to both ε and $\{g_0, \varphi_0\}$, while ε_{n+1} is the element with center ζ_{n+1} subordinate to both ε' and $\{g_n, \varphi_n\}$. Then for every $j = 0, 1, \ldots, n$, there is a chain of regular elements

$$\varepsilon_j = \varepsilon_0^{(j)}, \varepsilon_1^{(j)}, \ldots, \varepsilon_{k_j}^{(j)} = \varepsilon_{j+1}$$

with centers on L, all subordinate to $\{g_j, \varphi_j\}$, joining ε_j to ε_{j+1} (see Problem 8.6). Therefore the chain

$$\varepsilon, \varepsilon_0^{(0)}, \ldots, \varepsilon_{k_0}^{(0)} = \varepsilon_0^{(1)}, \ldots, \varepsilon_{k_1}^{(1)} = \varepsilon_0^{(2)}, \ldots, \varepsilon_{k_{n-1}}^{(n-1)} = \varepsilon_0^{(n)}, \ldots, \varepsilon_{k_n}^{(n)}, \varepsilon'$$

is a chain of regular elements, all subordinate to e_0, joining ε to ε' along L. as required.

COROLLARY. *Suppose* ε *and* ε' *have the same center. Then* ε' *is the continuation of* ε *along a circle* L *with center* z_0 *(traversed one or more times).*

Proof. In this case, τ and τ' lie on a circle with center at $t = 0$, and we can choose L to be an arc of this circle.

LEMMA 2. *Two generalized elements* e_0 *and* e' *with the same center coincide if they have a common subordinate element.*

Proof. The lemma is an immediate consequence of the interior uniqueness theorem if e_0 and e' are both unbranched. Thus suppose e_0, say, is a branched element of order $k - 1$. Then there are $k - 1$ other elements with the same center as ε which are subordinate to e_0 and distinct both from ε and from each other. According to Lemma 1 and its corollary, each of these elements is the result of continuation of ε along a circle with center z_0 (traversed several times in general), where each continuation can be achieved by using a chain of elements subordinate to e_0, all contained in the disk of e'. Considering each element of the chain in turn, we find that the result of each continuation is also subordinate to e', i.e., the $k - 1$ other elements concentric with ε and subordinate to e_0 are also subordinate to e'. Therefore e' is itself a branched element of order $\nu \geqslant k - 1$. Reversing the roles of e_0 and e' in this argument, we see that $\nu \leqslant k - 1$ and hence $\nu = k - 1$. Thus the mapping $t = (z - z_0)^{1/k}$ transforms e_0 and e' into two unbranched elements centered at $t = 0$, with a common subordinate element ε'. Therefore the transformed elements coincide, and hence so do the original elements.

LEMMA 3. *Any two elements* e_0 *and* e' *of an analytic configuration* A *can be joined by a chain of elements of* A *such that the elements other than* e_0 *and* e' *are all regular.*

Proof. The existence of a chain $e_0, e_1, \ldots, e_n = e'$ of elements of A joining e_0 to e' is an immediate consequence of the definition of A, but the intermediate elements $e_1, \ldots, e_{n-1}$ may not be regular. For example, suppose e_j $(1 \leqslant j \leqslant n - 1)$ is critical. Let ε_0 be an element subordinate to both e_{j-1} and e_j, and let ε' be an element subordinate to both e_j and e_{j+1}. These elements are both regular, and moreover, according to Lemma 1, they can be joined by a chain of regular elements subordinate to e_j. Replacing e_j by this chain, beginning with ε_0 and ending with ε', we obtain a new chain containing one less critical element than the original chain. Repeating this argument as many times as necessary, we arrive at a chain with no critical intermediate elements.

THEOREM 8.13. *The analytic configuration* A *is connected.*

Proof. Suppose A is disconnected. Then A is the union of two nonempty disjoint sets A_0 and A', neither of which contains an element which

is a "limit element" of the other (relative to the element topology introduced in the definition on p. 287). Given any two elements $e_0 \in A_0$, $e' \in A'$, there exist neighborhoods $\mathcal{N}(e_0) \subset A_0$, $\mathcal{N}(e') \subset A'$ (why?) and regular elements $\varepsilon_0 \in \mathcal{N}(e_0)$, $\varepsilon' \in \mathcal{N}(e')$. Then, according to Lemma 3, there is a chain $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n = \varepsilon'$ connecting ε_0 to ε' such that $\varepsilon_1, \ldots, \varepsilon_{n-1}$ are regular. Therefore every element of the chain is regular. Since $\varepsilon_0 \in A_0$, $\varepsilon' \in A'$, $A_0 \cap A' = 0$, there must exist two consecutive elements ε_{j-1} and ε_j, with disks K_{j-1} and K_j, such that $\varepsilon_{j-1} \in A_0$, $\varepsilon_j \in A'$. Let $g = K_{j-1} \cup K_j$, where g is a domain since $K_{j-1} \cap K_j \neq 0$. Given any point $\zeta \in g$, there is a unique element ε_ζ with center ζ which is subordinate to ε_{j-1} in K_{j-1}, to ε_j in K_j and to both ε_{j-1} and ε_j in $K_{j-1} \cap K_j$. Let g_0 and g' be the subsets of g such that $\varepsilon_\zeta \in A_0$ if $\zeta \in g_0$ while $\varepsilon_\zeta \in A'$ if $\zeta \in g'$. Then g_0 and g' are disjoint nonempty sets and $g_0 \cup g' = g$. Since g is connected, at least one of the sets g_0 and g' contains a limit point of the other. Suppose, for example, that $\alpha_0 \in g_0$ is a limit point of g'. Then the corresponding element of A_0 is a limit element of elements of A', contrary to hypothesis. This contradiction shows that A is connected.

COROLLARY. *The analytic configuration A is a two-dimensional manifold.*

THEOREM 8.14. *There are at most countably many elements of an analytic configuration A with the same center z_0.*

Proof. Suppose $z_0 = 0$ (otherwise make the preliminary transformation $\zeta = 1/z$ if $z_0 = \infty$ or $\zeta = z - z_0$ if $z_0 \neq 0, \infty$). Let $e_0 \in A$ be a fixed element with center 0 and $e \in A$ an arbitrary element with the same center. According to Lemma 3, there is a chain $e_0, e_1, \ldots, e_n = e$ of elements of A joining e_0 and e, with centers $z_0 = 0, z_1, \ldots, z_n$, where all the elements except possibly e_0 and e themselves are regular. It can be assumed without loss of generality that the centers z_1 and z_{n-1} are rational (i.e., have rational coordinates Re z_1, Im z_1, Re z_{n-1}, Im z_{n-1}) and that $e_1 \subset e_0$, $e_{n-1} \subset e_n$, since otherwise we need only insert extra elements $e_1' \subset e_0$, $e_{n-1}' \subset e_n$ with rational centers between the elements e_0 and e_1, or between e_{n-1} and e_n. Clearly, the element e_{n-1} is the result of continuing e_1 along the polygonal curve $\Lambda = \overline{z_1 z_2} + \cdots + \overline{z_{n-2} z_{n-1}}$. According to Theorem 8.4, we can replace Λ by another polygonal curve Λ' with rational vertices and the same end points such that the result of continuing e_1 along Λ' is still e_{n-1}. Since there are at most countably many polygonal curves with rational vertices and at most countably many elements with rational centers subordinate to a given element,[37] the number of distinct elements e_{n-1} obtained by continuations of this kind is at most countable. Every such e_{n-1} determines a unique element

[37] There are only finitely many elements with the same center subordinate to a given element.

e with center $z = 0$ such that $e_{n-1} \subset e$ (provided e exists). Hence there are at most countably many distinct elements e with center $z = 0$, and the theorem is proved.

THEOREM 8.15. *The analytic configuration A contains at most countably many critical elements (including branched elements).*

Proof. Let E be the set of all regular elements of A with rational centers. Then E is countable, since there are at most countably many such centers and at most countably many (regular) elements with the same center (by Theorem 8.14). To every critical element $e_0 \in A$ with center z_0 and radius R we assign a subordinate (regular) element $e \subset e_0$ with rational center ζ such that $|\zeta - z_0| < \frac{1}{2}R$. Since $\varepsilon \in E$, this maps the set of critical elements of A into the countable set E. Let e_0 and e' be distinct critical elements with disks K_0, K' and centers z_0, z', and suppose the regular element

$$\varepsilon: \sum_{n=0}^{\infty} a_n(z - \zeta)^n$$

is assigned (in the way described) to both e_0 and e', so that in particular the series of ε converges on the disks $\varkappa_0: |z - \zeta| < |\zeta - z_0|$ and $\varkappa': |z - \zeta| < |\zeta - z'|$ since $\varkappa_0 \subset K_0$, $z_0 \notin \varkappa_0$ and $\varkappa' \subset K'$, $z' \notin \varkappa'$ (e_0 has a pole or branch point at z_0, while e' has a pole or branch point at z'). Moreover, since $\varepsilon \subset e_0$, $\varepsilon \subset e'$, it is impossible for either of the quantities $|\zeta - z_0|$ and $|\zeta - z'|$ to exceed the other, and hence $|\zeta - z_0| = |\zeta - z'|$, $\varkappa_0 = \varkappa'$. Let $\gamma: |z - \zeta| = |z - z_0|$ be the common boundary of $\varkappa_0$ and $\varkappa'$. Then z_0 and z' both lie on γ and are singular points of ε (in the sense of Sec. I.84). But since $\gamma \subset K_0$, there can only be one singular point of $\varepsilon \subset e_0$ on γ, namely the point z_0. It follows that $z' = z_0$, and hence e_0 and e' have the same center. But then e_0 and e' coincide, by Lemma 2, since they have a common subordinate element ε. In other words, distinct critical elements of A are assigned distinct regular elements in the countable set E, and hence the set of all critical elements of A with finite centers is itself countable (possibly finite). To complete the proof, we note that according to Theorem 8.14, there are at most countably many critical elements centered at $z_0 = \infty$.

THEOREM 8.16. *The analytic configuration A has a countable base.*

Proof. Let E be the set consisting of all critical elements, all regular elements with rational centers and all regular elements with center $z_0 = \infty$, and let B be the set of all neighborhoods $\mathcal{N}(e; \rho)$, $e \in E$, where ρ is rational. Then E is countable (recall Theorems 8.14 and 8.15), and hence so is B. Moreover, B is a base for the analytic configuration A, in the sense of Definition 4, p. 218. In fact, consider any neighborhood $\mathcal{N}(e_0; \rho)$ where $e_0 \in A$. If e_0 is critical, then

$$e_0 \in \mathcal{N}(e_0; \rho) \subset \mathcal{N}(e_0; \rho_0) \tag{8.40}$$

for any rational $\rho \in (0, \rho_0)$. On the other hand, if e_0 is regular with center z_0, choose a regular element $e' \subset e_0$ with the rational center z' such that[38]

$$|z' - z_0| < \rho < \tfrac{1}{2}\rho_0,$$

where ρ is rational (the case $z_0 = \infty$ is handled by making the preliminary transformation $\zeta = 1/z$). Then the disk $|z - z'| < \rho$ contains z_0 and moreover is contained in the disk $|z - z_0| < \rho_0$. It follows that

$$e_0 \in \mathcal{N}(e'; \rho) \subset \mathcal{N}(e_0; \rho_0). \tag{8.41}$$

Comparing (8.40) and (8.41), we see that in any event there is a set $\mathcal{O} \in B$ such that $e_0 \in \mathcal{O} \subset \mathcal{N}(e_0; \rho_0)$, i.e., B is a base for A, as asserted.

Combining Theorems 8.12, 8.13 and 8.16, we obtain the central result of this section:

THEOREM 8.17. *The analytic configuration A is a topological surface.*

44. The Analytic Configuration as a Riemann Surface

Given an analytic configuration A, let $z = f(e)$ be the function mapping every element $e \in A$ into its center. Then $f(e)$ maps A into the Riemann sphere Σ (the extended z-plane),[39] and the point $e \in A$ or set $E \subset A$ is said to *lie over* its image $z = f(e)$ or $M = f(E)$. The function $f(e)$ is obviously continuous on A, since, given any $\rho > 0$, it follows from $z = f(e)$, $z_0 = f(e_0)$, $e \in \mathcal{N}(e_0; \rho)$ that $|z - z_0| < \rho$. If, in addition, $f(e)$ is homeomorphic on E, we say that E *lies univalently over* $M = f(E)$.

As we now proceed to show, the function $z = f(e)$ is actually an interior mapping relative to a suitable triangulation of A. Our first theorem asserts somewhat less, since it involves a collection of triangles $\mathcal{T}$ which in general is not a triangulation of A:

THEOREM 8.18. *Given an analytic configuration A, let $z = f(e)$ be the function mapping every element $e \in A$ into its center. Then there exists a collection of triangles $\mathcal{T}$ covering the whole surface A such that $f(e)$ and $\mathcal{T}$ have the two properties of an interior mapping given in the definition on p. 232.*

Proof. There are at most countably many branched elements of A lying over at most countably many points $\alpha_1, \ldots, \alpha_n, \ldots$ of the sphere Σ.

---

[38] It is tacitly assumed that ρ_0 does not exceed the radius of convergence of the series of e_0.

[39] Here we denote the points of Σ by z (instead of by w, as in Chap. 7).

Let $\{\tau_n\}$ be a sequence of triangulations of Σ such that

1. The triangulation τ_{n+1} is obtained from the triangulation τ_n by further subdivision of the triangles in τ_n;[40]
2. There are triangles in τ_n with α_n as a vertex ($n = 1, 2, \ldots$);
3. The quantity $|\tau_n|$, equal to the largest (set-theoretic) diameter of the triangles in τ_n, approaches zero as $n \to \infty$.

We now find every set of elements of A which lies univalently over a triangle in τ_1 and which contains either no branched elements at all or else a single branched element lying over the point α_1. Clearly every such set is a triangle Δ on the surface A, where Δ can contain a branched element only if the element is a vertex of Δ. Let $\mathscr{T}_1$ denote the set of all such triangles Δ. Next let $\mathscr{T}_2$ be the set of all triangles lying univalently over triangles in τ_2 which have not already been included in $\mathscr{T}_1$ and which contain either no branched elements at all or else a single branched element lying over one of the points α_1, α_2. Continuing the induction, we construct a sequence $\{\mathscr{T}_n\}$ of sets of triangles,[41] and then we form the union

$$\mathscr{T} = \mathscr{T}_1 \cup \cdots \cup \mathscr{T}_n \cup \cdots.$$

To show that $\mathscr{T}$ covers A, i.e., that every element $e_0 \in A$ belongs to at least one triangle in $\mathscr{T}$, we argue as follows: Suppose first that e_0 is an unbranched element with center z_0, and let $\mathscr{N}(e_0; \rho)$ be a neighborhood of e_0. Then, as on p. 287, $\mathscr{N}(e_0; \rho)$ is mapped homeomorphically by $z = f(e)$ onto the disk $K: |z - z_0| < \rho$. But $|\tau_n| < \rho$ for sufficiently large n, and hence $\delta \in \tau_n$, $z_0 \in \delta$ implies $\delta \subset K$. Therefore $f(e)$ maps every triangle $\Delta = f^{-1}(\delta) \subset \mathscr{N}(e_0; \rho)$ homeomorphically onto δ, i.e., Δ lies univalently over δ. Since Δ consists entirely of regular points, it must belong to the set $\mathscr{T}_n$, provided it has not already been included in one of the sets $\mathscr{T}_1, \ldots, \mathscr{T}_{n-1}$. Moreover, e_0 obviously belongs to Δ, by construction.

Next suppose that e_0 is a branched element of order k lying over the point α_m. It follows from conditions 1 and 2 that there are triangles in τ_n with vertex α_m provided that $n \geqslant m$. Let $\mathscr{N}(e_0; \rho)$ be a neighborhood of e_0. Then for sufficiently large n every triangle $\delta \in \tau_n$ with vertex α_m is contained in the disk $K: |z - z_0| < \rho$. The sum of the series of the element e_0 is k-valued and has k single-valued branches on int (δ). Corresponding to each of these branches, there is a distinct triangle Δ_j ($j = 1, \ldots, k$) lying univalently over δ which is mapped homeomorphically onto δ by the function $z = f(e)$. In other words, $\mathscr{N}(e_0; \rho)$ contains

[40] Explicitly, every vertex of a triangle in τ_n is a vertex of a triangle in τ_{n+1}, and every triangle in τ_{n+1} is contained in a triangle in τ_n (as usual, triangles are regarded as closed domains).

[41] Thus $\mathscr{T}_n$ consists of all triangles lying univalently over triangles of τ_n which are not already included in $\mathscr{T}_1, \ldots, \mathscr{T}_{n-1}$ and which contain either no branched elements at all or else a single branched element lying over one of the points $\alpha_1, \ldots, \alpha_n$.

k distinct triangles $\Delta_1, \ldots, \Delta_k$ lying univalently over δ, each containing one and only one branched element, namely e_0, at a common vertex. Moreover, the triangles $\Delta_1, \ldots, \Delta_k$ all belong to the set $\mathscr{T}_n$ provided they have not already been included in one of the sets $\mathscr{T}_1, \ldots, \mathscr{T}_{n-1}$.

Thus, finally, given any element $e_0 \in A$, we can find a triangle in $\mathscr{T}$ containing e_0. It follows that $\mathscr{T}$ covers the whole surface A. Moreover, the mapping $z = f(e)$ is homeomorphic on every triangle $\Delta \in \mathscr{T}$ and also in a neighborhood of every boundary point of Δ except possibly at a vertex (except for the vertices, the boundary points are all regular). Therefore $\mathscr{T}$ and $f(e)$ have the two properties figuring in the definition of an interior mapping, and the proof is complete.

Next we prove two lemmas, with the aim of eventually making $\mathscr{T}$ into a triangulation of A:

LEMMA 1. *Distinct triangles of $\mathscr{T}$ have disjoint interiors.*

Proof. Suppose to the contrary that two triangles $\Delta' \in \mathscr{T}_m$, $\Delta'' \in \mathscr{T}_n$ lying univalently over triangles $\delta' \in \tau_m$, $\delta'' \in \tau_n$ have a common interior point (element) ε, and suppose ε lies over the point ζ. Then the triangles δ' and δ'' have ζ as a common interior point, and hence must belong to different triangulations ($m \neq n$). Then one of these triangles must be contained in the other, say $\delta'' \subset \delta'$, which is possible only if $n > m$. The elements in Δ' and Δ'' define two single-valued analytic functions $\varphi_1(z)$ and $\varphi_2(z)$ on the simply connected domains int (δ_1) and int (δ_2). Since these functions coincide in a neighborhood of ζ, they must coincide on the whole triangle int (δ''), i.e., all the elements in int (Δ'') must be contained in int (Δ'), which implies $\Delta'' \subset \Delta'$. But this contradicts the definition of the set of triangles $\mathscr{T}_n$ (no triangle in $\mathscr{T}_n$ can be a triangle of $\mathscr{T}_m$ if $n > m$), thereby establishing the lemma.

LEMMA 2. *Every element $e_0 \in A$ has a neighborhood covered by finitely many triangles of $\mathscr{T}$.*

Proof. In proving the existence of triangles in $\mathscr{T}$ containing any given element $e_0 \in A$, we found that for sufficiently large n there is a neighborhood of e_0 covered by finitely many triangles $\Delta_1, \ldots, \Delta_n \in \mathscr{T}$. But there can be no other triangles in $\mathscr{T}$ participating in this covering, since any such triangle must share an interior point with one of the triangles $\Delta_1, \ldots, \Delta_n$ and hence, by Lemma 1, must coincide with one of these triangles.

The final step in proving that A is a Riemann surface can now be taken:

THEOREM 8.19. *Given an analytic configuration A, let $z = f(e)$ be the function mapping every element $e \in A$ into its center. Then there exists a triangulation of A relative to which $f(e)$ is an interior mapping. In particular, A is a Riemann covering surface.*

Proof. Consider any triangle Δ belonging to the set $\mathscr{T}$ constructed in the proof of Theorem 8.18, and suppose Δ lies over the triangle $\delta \subset \Sigma$. Then, by Lemma 2, the boundary of Δ can contain only finitely many vertices of triangles of $\mathscr{T}$. Suppose these vertices lie over the boundary points $\zeta_1, \ldots, \zeta_n$ of the triangle δ. Let ζ be any interior point of δ, and draw Jordan arcs joining ζ to all the points $\zeta_1, \ldots, \zeta_n$, making sure that the arcs are disjoint except for the point ζ and that none of the arcs intersects the boundary of δ in a point other than its end point ζ_j $(j = 1, \ldots, n)$. As a result, δ is divided into n triangles $\delta_1, \ldots, \delta_n$ with common vertex ζ, and correspondingly, Δ is divided into n triangles $\Delta_1, \ldots, \Delta_n$ lying over the triangles $\delta_1, \ldots, \delta_n$. Carrying out a similar construction for all the triangles in $\mathscr{T}$, of which there are only countably many (why?), we finally obtain a triangulation of the surface A. Moreover, if Δ is any triangle in the resulting triangulation, the mapping $z = f(e)$ is homeomorphic on Δ and also in a neighborhood of every boundary point of Δ distinct from a vertex (the branched elements all lie at vertices). Therefore $f(e)$ is an interior mapping, as asserted.

45. Algebraic Functions

By an *algebraic function* we mean a solution $w = f(z)$ of an algebraic equation of the form

$$F(z, w) \equiv p_0(z) + p_1(z)w + \cdots + p_n(z)w^n = 0, \tag{8.42}$$

where it is assumed that

1. Every $p_j(z)$ is a polynomial in z;
2. $p_n(z) \not\equiv 0$, so that (8.42) is of degree n in w;
3. The polynomials $p_0(z), p_1(z), \ldots, p_n(z)$ have no common factors of degree greater than 0 and hence no common zeros.

We also assume that equation (8.42) is *irreducible*, i.e., that the polynomial $F(z, w)$ cannot be represented as a product

$$F(z, w) = F_1(z, w)F_2(z, w),$$

where $F_1(z, w)$ and $F_2(z, w)$ are polynomials of degree less than n in w. Otherwise, (8.42) could be replaced by a pair of equations

$$F_1(z, w) = 0, \qquad F_2(z, w) = 0, \tag{8.43}$$

and determination of w as a function of z would then reduce to investigating each of the equations (8.43) separately.

Let $z_1, \ldots, z_\mu$ be the distinct points which are zeros of the polynomial $p_n(z)$ or of the discriminant $D(z)$ of equation (8.42) [see p. II.110]. Then at

every finite point z_0 other than the points $z_1, \ldots, z_\mu$, equation (8.42) has n distinct finite roots $w_1, \ldots, w_n$ satisfying the conditions

$$F(z_0, w_j) = 0, \quad \frac{\partial F(z_0, w_j)}{\partial w} \neq 0 \qquad (j = 1, \ldots, n).$$

It follows from the implicit function theorem (Theorem II.3.11) that z_0 has a neighborhood $\mathcal{N}(z_0)$ in which there can be defined n (and only n) distinct single-valued analytic functions $f_1(z), \ldots, f_n(z)$, called *local solutions* of (8.42), such that

$$f_j(z_0) = w_j, \quad F[z, f_j(z)] = 0 \qquad (j = 1, \ldots, n). \tag{8.44}$$

In other words, at every finite point z_0 other than $z_1, \ldots, z_\mu$ there are precisely n regular elements $f_j(z)$,[42] all satisfying equation (8.42).

THEOREM 8.20. *Let G be the domain obtained by deleting the points $z_1, \ldots, z_\mu$ from the finite z-plane. Then each of the n elements $f_1(z), \ldots, f_n(z)$ with center $z_0 \in G$ obtained by solving* (8.42) *locally can be continued analytically along any curve $L \subset G$.*

Proof. Let

$$L: z = \lambda(t) \qquad (\alpha \leqslant t \leqslant \beta)$$

be a continuous curve contained in G, with initial point z_0, and let $f_{j_0}(z)$ be an element with center z_0. Suppose $\zeta = \lambda(\tau)$, $\alpha < \tau \leqslant \beta$ is the first singular point encountered in continuing $f_{j_0}(z)$ analytically along L. Then, as just shown, ζ has a neighborhood $\mathcal{N}(\zeta)$ in which there are precisely n regular elements $\varphi_1(z), \ldots, \varphi_n(z)$ satisfying equation (8.42). Let $\zeta_1 = \lambda(t_1)$ be a point of the curve L such that

$$t_1 < \tau, \quad \zeta_1 \in \mathcal{N}(\zeta), \qquad L[t_1, \tau] \subset \mathcal{N}(\zeta),$$

and let $\psi_1(z), \ldots, \psi_n(z)$ be the elements with center ζ_1 which are subordinate to $\varphi_1(z), \ldots, \varphi_n(z)$. Since the elements $\varphi_j(z)$ satisfy equation (8.42) and are distinct, the same is true of the elements $\psi_j(z)$. Let $\chi(z)$ be the element with center ζ_1 obtained from $f_{j_0}(z)$ by continuation along $L[\alpha, t_1]$. Then $\chi(z)$ must satisfy (8.42) with $z = \zeta_1$ (why?), and hence must coincide with one of the elements $\psi_1(z), \ldots, \psi_n(z)$, say $\psi_{k_0}(z)$, since the $\psi_k(z)$ are the only solutions of (8.42) for $z = \zeta_1$. But $\psi_{k_0}(z)$ is subordinate to the element $\varphi_{k_0}(z)$ with center ζ, and the disks of these two elements cover the arc $L[t_1, \tau]$. It follows that $f_{j_0}(z)$ can be continued along the whole arc $L[\alpha, \tau]$, contrary to the assumption that ζ is a singular point. Therefore L contains no singular points, i.e., $f_{j_0}(z)$ can be continued along L as asserted, and in fact the continuation is accomplished by using elements satisfying the same equation (8.42) as $f_{j_0}(z)$ itself.

---

[42] More exactly, n regular elements $\{\mathcal{N}(z_0), f_j(z)\}$.

COROLLARY. *Let $\mathcal{N}' = \mathcal{N}'(z_j; \rho)$ be a deleted neighborhood of z_j containing none of the exceptional points $z_1, \ldots, z_\mu$ other than z_j itself, and let $f_{j_0}(z)$ be any of the elements with center $z_0 \in \mathcal{N}'$ obtained by solving* (8.42) *locally. Then the result of continuing $f_{j_0}(z)$ analytically along all curves $L \subset \mathcal{N}'$ is a solution $f(z)$ of* (8.42) *with the representation*

$$f(z) = \sum_{l=-\infty}^{\infty} a_l(z - z_j)^{l/k} \tag{8.45}$$

for all $z \in \mathcal{N}'$, where k is a positive integer not exceeding n.

Proof. The function $f(z)$ can be single-valued or multiple-valued on $\mathcal{N}'$, but not more than n-valued. Correspondingly, it is possible *a priori* for z_j to be a regular point, a pole or an essential singular point if $f(z)$ is single-valued, or an algebraic or transcendental branch point (of order not exceeding $n - 1$) if $f(z)$ is multiple-valued (however, see Theorem 8.21 below). These various possibilities are all included in the representation (8.45).

LEMMA. *Given a polynomial*

$$P(w) = c_0 + c_1 w + \cdots + c_{n-1}w^{n-1} + w^n,$$

let $M > 0$ be such that

$$|c_j| \leqslant M \qquad (j = 0, 1, \ldots, n - 1).$$

Then the zeros of $P(w)$ depend continuously on the coefficients of $P(w)$ in the following sense: Given any $\varepsilon > 0$ and any k $(1 \leqslant k \leqslant n)$, there is a $\delta = \delta(\varepsilon, M, k)$ such that $P(w)$ has at least k zeros of absolute value less than ε if

$$|c_j| < \delta \qquad (j = 0, 1, \ldots, k - 1).$$

Proof. Let $w_1, \ldots, w_n$ be the zeros of $P(w)$ arranged in order of increasing absolute value:

$$|w_1| \leqslant |w_2| \leqslant \cdots \leqslant |w_n|.$$

If $|w| > M + 1$, then

$$|P(w)| \geqslant |w|^n\left(1 - \frac{M}{|w|} - \cdots - \frac{M}{|w|^n}\right) > |w|^n\left(1 - \frac{M}{|w| - 1}\right) > 0,$$

and hence no zero has absolute value greater than $M + 1 = M'$. Noting that

$$c_0 = (-1)^n w_1 \cdots w_n,$$

we find that

$$|w_1| \leqslant |w_1 \cdots w_n|^{1/n} = |c_0|^{1/n} < \varepsilon$$

if

$$|c_0| < \delta_1 = \varepsilon^n.$$

Suppose it has been shown that

$$|w_j| < \varepsilon \qquad (j = 0, 1, \ldots, m < k)$$

provided that

$$|c_j| < \delta_m(\varepsilon, M) \qquad (j = 0, 1, \ldots, m - 1). \tag{8.46}$$

Except for sign, c_m equals the sum of all possible products of zeros of $P(w)$ with different indices taken $n - m$ at a time, and hence

$$|c_m| \geqslant |w_{m+1} \cdots w_n| - \Sigma |w_{j_1} \cdots w_{j_{n-m}}|,$$

where the sum contains $\binom{n}{m} - 1$ terms each involving at least one factor with an index not exceeding m. Therefore (8.46) implies

$$|w_{m+1}| \leqslant |w_{m+1} \cdots w_n|^{\frac{1}{n-m}} \leqslant \left\{ |c_m| + \left[\binom{n}{m} - 1 \right] M'^{n-m-1} \varepsilon \right\}^{\frac{1}{n-m}}$$

$$< M' \binom{n}{m}^{\frac{1}{n-m}} \varepsilon^{\frac{1}{n-m}} = \varepsilon'$$

if $|c_m| < M'^{n-m-1}\varepsilon$. Setting

$$\delta_{m+1} = \min (\delta_m, M'^{n-m-1}\varepsilon),$$

we find that

$$|w_j| < \varepsilon' \qquad (j = 0, 1, \ldots, m + 1)$$

provided that

$$|c_j| < \delta_{m+1}(\varepsilon, M) \qquad (j = 0, 1, \ldots, m).$$

The lemma now follows at once by induction.

THEOREM 8.21. *All but a finite number of coefficients a_l with negative l vanish in* (8.45), *i.e., z_j cannot be an essential singular point or a transcendental branch point of $f(z)$.*

Proof. The proof is sufficiently complicated to warrant being divided into three steps:

Step 1. First suppose that $p_n(z_j) \neq 0$. Then the equation

$$F(z_j, w) = p_0(z_j) + p_1(z_j)w + \cdots + p_n(z_j)w^n = 0$$

has roots $\omega_1, \ldots, \omega_r$ with multiplicities $\alpha_1, \ldots, \alpha_r$ where

$$\sum_{q=1}^{r} \alpha_q = n. \tag{8.47}$$

Writing $w = \omega_q + w'$, we have

$$F(z, w) = \pi_0(z) + \pi_1(z)w' + \cdots + \pi_n(z)w'^n,$$

where $\pi_n(z) = p_n(z)$ and hence $\pi_n(z) \neq 0$ in some neighborhood U_0 of the point z_j. For $z \in U_0$ the equation $F(z, w) = 0$ can be written in the form

$$\frac{\pi_0(z)}{\pi_n(z)} + \frac{\pi_1(z)}{\pi_n(z)} w' + \cdots + \frac{\pi_{n-1}(z)}{\pi_n(z)} w'^{n-1} + w'^n = 0. \qquad (8.48)$$

Since ω_q is a root of multiplicity α_q of the equation $F(z_j, w) = 0$, (8.48) has a root of multiplicity α_q at the point $w' = 0$. Therefore the coefficients

$$\frac{\pi_s(z)}{\pi_n(z)} \qquad (s = 0, 1, \ldots, \alpha_q - 1)$$

vanish for $z = z_j$.

We now apply the lemma to equation (8.48). Making $\varepsilon > 0$ so small that the neighborhoods

$$V_1 = \mathcal{N}(\omega_1; \varepsilon), \ldots, V_r = \mathcal{N}(\omega_r; \varepsilon)$$

in the w-plane are disjoint, we choose $\delta_q = \delta(\varepsilon, M, \alpha_q)$ where

$$M = \max_{1 \leq s \leq \alpha_q - 1} \sup_{z \in U_0} \left| \frac{\pi_s(z)}{\pi_n(z)} \right|.$$

Using the continuity of the rational functions $\pi_s(z)/\pi_n(z)$ on U_0, we can replace U_0 by a smaller neighborhood U_q of the point z_j such that

$$\left| \frac{\pi_s(z)}{\pi_n(z)} \right| < \delta_q \qquad (s = 0, 1, \ldots, \alpha_q - 1)$$

for all $z \in U_q$. We then find that for every $z \in U_q$, (8.48) has at least α_q roots (when solved for w') in the neighborhood $|w'| < \varepsilon$, and hence the equation $F(z, w) = 0$ has at least α_q roots (when solved for w) in the neighborhood V_q.

Now let U be any neighborhood of the point z_j contained in all the neighborhoods U_q $(q = 1, \ldots, r)$. Then for every $z \in U$, the equation $F(z, w) = 0$ has at least α_q roots in V_q $(q = 1, \ldots, r)$. Because of (8.47) and the fact that the neighborhoods $V_1, \ldots, V_r$ are disjoint, all the roots of $F(z, w) = 0$ are contained in the neighborhoods $V_1, \ldots, V_r$, and each V_q contains precisely α_q roots.

Turning to the function $f(z)$ figuring in (8.45), we choose ρ so small that the deleted neighborhood $\mathcal{N}' = \mathcal{N}'(z_j; \rho)$ is contained in U. Since the values of $f(z)$ are all roots of the equation $F(z, w) = 0$, they must all belong to the disjoint neighborhoods $V_1, \ldots, V_r$. On the other hand, $f(z)$ carries $\mathcal{N}'$ into a connected set, in fact a domain (why?), and hence $f(\mathcal{N}')$ must lie entirely in one of the neighborhoods $V_1, \ldots, V_r$, say V_{q_0}. In other words, given any $\varepsilon > 0$, there is a number $\rho = \rho(\varepsilon) > 0$ such that $|f(z) - \omega_{q_0}| < \varepsilon$ whenever $0 < |z - z_j| < \rho$. Therefore all

the coefficients a_l with negative l in the expansion (8.45) must vanish, since otherwise $f(z)$ would approach infinity as $z \to z_j$. Thus, finally, we have the representation

$$f(z) = \sum_{l=0}^{\infty} a_l(z - z_j)^{l/k},$$

i.e., if $p_n(z_j) \neq 0$, then z_j must either be a regular point of $f(z)$ [if $k = 1$] or an algebraic branch point. At the point z_j itself, we define

$$f(z_j) = \lim_{z \to z_j} f(z) = \omega_{q_0} = a_0.$$

Step 2. Next suppose that $p_n(z_j) = 0$, so that

$$p_n(z) = (z - z_j)^{\alpha} P_n(z),$$

where α is a positive integer and $P_n(z)$ is a polynomial which does not vanish at z_j. Then the equation $F(z, w) = 0$ can be replaced by

$$(z - z_j)^{\alpha(n-1)} F(z, w) = p_0(z)(z - z_j)^{\alpha(n-1)} + p_1(z)(z - z_j)^{\alpha(n-1)} w + \cdots + P_n(z)(z - z_j)^{\alpha n} w^n = 0,$$

or

$$P_0(z) + P_1(z)W + \cdots + P_n(z)W^n = 0 \tag{8.49}$$

in terms of the new variable

$$W = (z - z_j)^{\alpha} w$$

and the polynomials

$$P_j(z) = p_j(z)(z - z_j)^{\alpha(n-1-j)} \qquad (j = 1, \ldots, n-1).$$

Equation (8.49) is satisfied in $\mathcal{N}'$ by the function

$$W = (z - z_j)^{\alpha} f(z),$$

and hence by the same argument as in Step 1,

$$(z - z_j)^{\alpha} f(z) = \sum_{l=0}^{\infty} a_l'(z - z_j)^{l/k},$$

since $P_n(z_j) \neq 0$. It follows that

$$f(z) = \sum_{l=0}^{\infty} a_l'(z - z_j)^{(l/k)-\alpha} = \sum_{l=-\alpha k}^{\infty} a_l(z - z_j)^{l/k}.$$

Therefore in this case all but a finite number of the coefficients a_l with negative l in the expansion (8.45) vanish, i.e., z_j is either a pole of $f(z)$ (if $k = 1$) or an algebraic branch point. Thus $f(z)$ now approaches ∞ as $z \to z_j$, unless all the coefficients a_l with negative l vanish (in which case z_j is a regular point if $k = 1$).

Step 3. Finally let $\mathcal{N}'$ be a deleted neighborhood of infinity containing none of the points $z_1, \ldots, z_\mu$, and consider the function $f(z)$ obtained

by continuing any element $f_{z_0}(z)$ with center $z_0 \in \mathcal{N}'$ satisfying (8.42) along all curves $L \subset \mathcal{N}'$. The substitution $\zeta = 1/z$ transforms (8.42) into

$$p_0\left(\frac{1}{\zeta}\right) + p_1\left(\frac{1}{\zeta}\right)w + \cdots + p_n\left(\frac{1}{\zeta}\right)w^n = 0,$$

which after multiplication by a suitable power of ζ takes the form

$$q_0(\zeta) + q_1(\zeta)w + \cdots + q_n(\zeta)w^n = 0,$$

where the $q_j(z)$ are polynomials in ζ. Clearly this equation must be satisfied by the function $f(1/\zeta)$ in a deleted neighborhood $\tilde{\mathcal{N}}'$ of the point $\zeta = 0$. Therefore, by Steps 1 and 2,

$$f\left(\frac{1}{\zeta}\right) = \sum_{l=0}^{\infty} a_l^{l/k} \qquad (\zeta \in \tilde{\mathcal{N}}')$$

if $q_n(0) \neq 0$, while

$$f\left(\frac{1}{\zeta}\right) = \sum_{l=m}^{\infty} a_l^{l/k} \qquad (\zeta \in \tilde{\mathcal{N}}',\ m \leqslant 0)$$

if $q_n(0) = 0$. Returning to the variable $z = 1/\zeta$, we have

$$f(z) = \sum_{l=0}^{\infty} a_l z^{-l/k} \qquad (z \in \mathcal{N}')$$

or

$$f(z) = \sum_{l=m}^{\infty} a_l z^{-l/k} \qquad (z \in \mathcal{N}',\ m \leqslant 0).$$

Thus the point $z = \infty$ cannot be an essential singular point or a transcendental branch point of $f(z)$, and we again set

$$f(\infty) = \lim_{z \to \infty} f(z).$$

The proof of Theorem 8.21 is now complete.

COROLLARY. *The solution of the algebraic equation* (8.42) *of degree n is an n-valued function* $w = f(z)$ *defined in the extended z-plane. The function* $w = f(z)$ *is represented by elements of the form*

$$\sum_{l=m}^{\infty} a_l(z - z_0)^{l/k} \qquad (1 \leqslant k \leqslant n,\ m \leqslant 0)$$

if z_0 *is finite and equal to one of the zeros* $z_1, \ldots, z_\mu$ *of the polynomial* $p_n(z)$ *or of the discriminant* $D(z)$ *of* (8.42), *and by an element of the form*

$$\sum_{l=m}^{\infty} a_l z^{-l/k} \qquad (1 \leqslant k \leqslant n,\ m \leqslant 0)$$

if $z_0 = \infty$. *If* $z_0 \neq z_1, \ldots, z_\mu, \infty$, *then* $w = f(z)$ *is represented by* n *regular elements with center* z_0.

THEOREM 8.22. *Let* A *be the set of all elements figuring in the corollary to Theorem* 8.21. *Then* A *is an analytic configuration.*

Proof. We must prove that any element in A is the result of analytic continuation of any other element in A. It is sufficient to show that any regular element satisfying the equation $F(z, w) = 0$ can be continued into any other regular element with the same center satisfying $F(z, w) = 0$.[43] In fact, let $f_{j_0}(z)$ and $\varphi_{j_1}(z)$ be two elements in A with distinct centers z_0 and ζ. Then, using Theorem 8.20, we can first continue $f_{j_0}(z)$ into an element $\varphi_{j_0}(z)$ with center ζ, and afterwards continue $\varphi_{j_0}(z)$ into the element $\varphi_{j_1}(z)$. The case of critical elements satisfying $F(z, w) = 0$ need not be considered, since they were originally *defined* by analytic continuation of regular elements (see the corollary to Theorem 8.20).

The fact that any regular element can be continued into any other regular element with the same center will be established by contradiction. Thus suppose an element $f_1(z)$ with center z_0 can be continued into the elements

$$f_2(z), \ldots, f_\nu(z) \qquad (\nu < n)$$

with the same center, but not into the remaining elements

$$f_{\nu+1}(z), \ldots, f_n(z).$$

Continuing $f_1(z)$ along all possible paths in the domain G (the finite plane minus the exceptional points $z_1, \ldots, z_\mu$), we obtain a function $\tilde{f}(z)$ which is in general ν-valued. Given any closed curve $L \subset G$ beginning and ending at z_0, let $f_{\alpha_1}(z)$ be the result of continuing $f_1(z)$ along L in the positive direction. If $f_{\alpha_1}(z)$ is distinct from $f_1(z)$, we continue $f_{\alpha_1}(z)$ along L in the same direction, obtaining a new element $f_{\alpha_2}(z)$ distinct from $f_{\alpha_1}(z)$.[44] If $f_{\alpha_2}(z)$ is also distinct from $f_1(z)$, we continue $f_{\alpha_2}(z)$ along L again, obtaining an element $f_{\alpha_3}(z)$ distinct from $f_{\alpha_1}(z)$ and $f_{\alpha_2}(z)$. Repeating this process, we eventually obtain an element $f_{\alpha_{l-1}}(z)$ whose continuation along L is the original element $f_1(z)$. In fact, there are at most ν distinct elements with center z_0 which can be obtained by all possible continuations of $f_1(z)$, and hence the above process of generating elements $f_{\alpha_1}(z), f_{\alpha_2}(z), \ldots$ distinct from $f_1(z)$ and from each other must

[43] If an element e_1 is the result of continuation of another element e_0 (along some curve L), we say that e_0 can be "continued into" e_1 (and *vice versa*).

[44] Otherwise the result of continuing $f_{\alpha_1}(z)$ along L in the negative direction would again be $f_{\alpha_1}(z)$ instead of $f_1(z)$. More generally, continuation of $\lambda \leqslant \nu$ distinct elements $f_1(z), \ldots, f_\lambda(z)$ along any curve L gives λ distinct elements $\varphi_1(z), \ldots, \varphi_\lambda(z)$, since if $\varphi_s(z) = \varphi_t(z) = \varphi(z)$, then continuation of $\varphi(z)$ along L in the negative direction would imply $f_s(t) = f_t(z)$, contrary to hypothesis.

terminate after at most $\nu - 1$ steps. Thus to every closed curve L there corresponds a *cycle* of j elements[45]

$$f_1(z), \quad f_{\alpha_1}(z), \ldots, \quad f_{\alpha_{j-1}}(z) \tag{8.50}$$

such that each element is obtained from the preceding element by continuation along L in the positive direction, except for $f_1(z)$ which is the result of continuing $f_{\alpha_{j-1}}(z)$ along L. If $j < \nu$, we choose an element $f_{\beta_0}(z)$ distinct from the elements (8.50) and continue $f_{\beta_0}(z)$ repeatedly along L, thereby obtaining a new cycle

$$f_{\beta_0}(z), \quad f_{\beta_1}(z), \ldots, \quad f_{\beta_{j'-1}}(z)$$

of elements distinct from the elements (8.50). In this way, starting from the elements $f_1(z), \ldots, f_\nu(z)$, we construct a set of cycles (relative to a given closed curve $L \subset G$) such that no two cycles have an element in common.

We now form the product

$$[w - f_1(z)] \cdots [w - f_\nu(z)] = w^\nu + A_{\nu-1}w^{\nu-1} + \cdots + A_0(z). \tag{8.51}$$

Since the elements $f_1(z), \ldots, f_\nu(z)$ can be continued analytically along any curve $L \subset G$, the same is true of the coefficients $A_0(z), \ldots, A_{\nu-1}(z)$, which, apart from sign, are the elementary symmetric functions in $f_1(z), \ldots, f_\nu(z)$.[46] Therefore the functions $A_0(z), \ldots, A_{\nu-1}(z)$ are analytic (and *a priori* ν-valued) on G. However, continuation of the set of elements $f_1(z), \ldots, f_\nu(z)$ with center z_0 along any closed curve $L \subset G$ leads to the same set of elements, arranged in a different order which can be deduced from a knowledge of the cycles corresponding to L. Therefore continuation of any element of the symmetric function $A_s(z)$, $s = 0, 1, \ldots, \nu - 1$ along any closed curve $L \subset G$ leads to the same element again, i.e., $A_s(z)$ is single-valued on G. In particular, every point $z_1, \ldots, z_\mu, \infty$ in the boundary of G is a possible regular point, pole or essential singular point of $A_s(z)$. However, the case of an essential singular point is excluded. In fact, the multiple-valued function $f(z)$ has a representation of the form

$$\sum_{l=m}^{\infty} a_l(z - z_j)^{l/\nu}$$

in a deleted neighborhood of every finite boundary point z_j $(j = 1, \ldots, k)$ and a representation of the form

$$\sum_{l=m}^{\infty} a_l z^{-l/\nu}$$

---

[45] If $j = 0$, (8.50) reduces to the single element $f_1(z)$.

[46] See e.g., G. Birkhoff and S. MacLane, *op. cit.*, Chap. 6, Sec. 10. Also recall p. II.106.

in a deleted neighborhood of ∞ (recall the corollary to Theorem 8.21). Substituting the corresponding series for the elements $f_1(z), \ldots, f_\nu(z)$ into the expressions for the symmetric functions $A_0(z), \ldots, A_{\nu-1}(z)$, and using the absolute convergence of the series to carry out the required operations of multiplication and addition, we find that every $A_s(z)$ has an expansion at z_j $(j = 1, \ldots, \mu)$ containing only finitely many terms with negative powers of $z - z_j$ and an expansion at ∞ containing only finitely many terms with positive powers of z. In other words, none of the functions $A_0(z), \ldots, A_{\nu-1}(z)$ can have an essential singular point in the extended plane. Therefore, by Theorem II.10.9, every $A_s(z)$ is a rational function and hence can be written as a fraction

$$A_s(z) = \frac{\tilde{P}_s(z)}{Q_s(z)} \qquad (s = 0, 1, \ldots, \nu - 1)$$

in lowest terms, where $\tilde{P}_s(z)$ and $Q_s(z)$ are polynomials.[47]

Denoting the least common multiple of the polynomials $Q_1(z), \ldots, Q_{\nu-1}(z)$ by $Q(z)$ $[Q(z) \not\equiv 0]$, we can write (8.51) in the form

$$[w - f_1(z)] \cdots [w - f_\nu(z)] = \frac{1}{Q(z)} [P_\nu(z)w^\nu + P_{\nu-1}(z)w^{\nu-1} + \cdots + P_0(z)],$$

where

$$P_\nu(z) = Q(z), \quad P_s(z) = \frac{\tilde{P}_s(z)Q(z)}{Q_s(z)} \qquad (s = 0, 1, \ldots, \nu - 1).$$

Therefore the ν-valued function $\tilde{f}(z)$ resulting from all possible continuations of an element $f_1(z)$ satisfying the algebraic equation (8.42) of degree n is itself the solution of an algebraic equation

$$P_\nu(z)w^\nu + P_{\nu-1}(z)w^{\nu-1} + \cdots + P_0(z) = 0 \tag{8.52}$$

of degree $\nu < n$. The coefficients of this equation have no common factors of degree greater than 0 (i.e., no common zeros). In fact, suppose $z - a$ divides the coefficient $P_\nu(z) = Q(z)$. Then $z - a$ divides at least one denominator of the fractions $\tilde{P}_s(z)/Q_s(z)$. Suppose $Q_{s_0}(z)$ contains the factor $z - a$ raised to the highest power. Then the corresponding numerator $\tilde{P}_{s_0}(z)$ is not divisible by $z - a$ and moreover neither is $Q(z)/Q_{s_0}(z)$. It follows that

$$P_{s_0}(z) = \frac{Q(z)}{Q_{s_0}(z)} \tilde{P}_{s_0}(z)$$

is not divisible by $z - a$, and hence $z - a$ cannot be a common factor of the coefficients of (8.52).

---

[47] If $A_s(z) \equiv 0$, we set $\tilde{P}_s(z) \equiv 0$ and $Q_s(z) \equiv 1$.

Now consider the polynomial

$$F(z, w) = p_0(z) + \cdots + p_n(z)w^n$$

figuring in the original equation (8.42), and also the polynomial

$$\Phi(z, w) = P_0(z) + \cdots + P_\nu(z)w^\nu.$$

The factorizations

$$F(z, w) = p_n(z)[w - f_1(z)]\cdots[w - f_n(z)],$$
$$\Phi(z, w) = P_\nu(z)[w - f_1(z)]\cdots[w - f_\nu(z)]$$

hold for all possible continuations of the regular elements $f_1(z), \ldots, f_n(z)$ along curves $L \subset G$, and hence

$$\frac{F(z, w)}{\Phi(z, w)} = \frac{p_n(z)}{P_\nu(z)}[w - f_{\nu+1}(z)]\cdots[w - f_n(z)].$$

But everything proved above for the product $[w - f_1(z)]\cdots[w - f_\nu(z)]$ is equally true of the product $[w - f_{\nu+1}(z)]\cdots[w - f_n(z)]$, and hence

$$[w - f_{\nu+1}(z)]\cdots[w - f_n(z)] = \frac{1}{R_{n-\nu}(z)}[R_{n-\nu}(z)w^{n-\nu} + \cdots + R_0(z)],$$

where $R_0(z), \ldots, R_{n-\nu}(z)$ are polynomials with no common zeros. Therefore

$$\frac{F(z, w)}{\Phi(z, w)} = \frac{p_n(z)}{P_\nu(z)R_{n-\nu}(z)}[P_0(z) + \cdots + P_\nu(z)w^\nu] \times [R_0(z) + \cdots + R_{n-\nu}(z)w^{n-\nu}]. \tag{8.53}$$

Suppose that

$$\frac{p_n(z)}{P_\nu(z)R_{n-\nu}(z)} = \frac{\hat{P}(z)}{\hat{Q}(z)},$$

where the fraction on the right is in lowest terms. Clearly the polynomial $\hat{P}(z)$ is of degree 0, since otherwise $\hat{P}(z)$ would have at least one zero $z = a$ and then (8.53) would imply $F(a, w) = 0$ for all w, which is impossible since the coefficients $p_0(a), \ldots, p_n(a)$ of $F(a, w)$, regarded as a polynomial in w, cannot all vanish. Similarly, we see that the polynomial $\hat{Q}(z)$ is of degree 0, since otherwise $\hat{Q}(z)$ would have at least zero $z = b$ and then (8.53) would imply $F(b, w) = \infty$ for all w, which is impossible.[48]

Thus (8.53) finally takes the form

$$F(z, w) = C[P_0(z) + \cdots + P_\nu(z)w^\nu][R_0(z) + \cdots + R_{n-\nu}(z)w^{n-\nu}],$$

[48] In asserting that the right-hand side of (8.53) becomes infinite for $z = b$, we use the fact that the polynomials $P_0(z), \ldots, P_\nu(z)$ have no common zeros, and similarly for the polynomials $R_0(z), \ldots, R_{n-\nu}(z)$.

where C is a nonzero constant. But this is impossible, since the original algebraic equation (8.42) is irreducible. Therefore ν cannot be less than n, i.e., the set A of all elements satisfying (8.42) is an analytic configuration, as asserted.

To recapitulate, the solution of the irreducible algebraic equation (8.42) is a multiple-valued function $w = f(z)$ with n single-valued analytic branches on the domain G obtained by deleting the points $z_1, \ldots, z_\mu, \infty$ from the extended z-plane. The function $f(z)$ can have singular points at the boundary points $z_1, \ldots, z_\mu, \infty$, but these can only be poles or algebraic branch points. We call $f(z)$ an *algebraic function*, and the corresponding analytic configuration A, regarded as a Riemann covering surface, is called the *Riemann surface* of $f(z)$. This surface is n-sheeted, since A contains n distinct regular elements lying over every point of G. The branched elements of A can only lie over the points $z_1, \ldots, z_\mu, \infty$ and correspond to branch points of the Riemann surface.

To construct a topological model of the Riemann surface of $f(z)$, we start from n "replicas" of the extended z-plane, each cut along a polygonal Jordan curve Λ with μ segments, joining the points $z_1, \ldots, z_\mu, \infty$ in some order (see Problem 8.25). We then paste these replicas together along the various "cuts" (segments of Λ) according to the following prescription: Given n distinct regular elements $e_1, \ldots, e_n$ of A with the same center $z_0 \in G$, $z_0 \in \Lambda$, we continue each element into one of the replicas, thereby generating n single-valued analytic branches of $f(z)$.[49] The branch obtained by continuation of e_j ($j = 1, \ldots, n$) will be called the "jth branch," defined on the "jth replica." If e_j is continued along a curve terminating at a point $\zeta \in \Lambda$ ($\zeta \neq z_0, z_1, \ldots, z_\mu, \infty$), the result is another regular element of A, which we denote by e_j^+ or e_j^- depending on whether ζ is approached from above or below the cut containing ζ. If $e_j^+ = e_j^-$, we paste the upper and lower edges of the cut together in the jth replica. This has the effect of "erasing" the cut from the jth replica. If $e_j^+ \neq e_j^-$, we find the unique value of l ($1 \leqslant l \leqslant n$) such that $e_j^+ = e_l^-$, and then paste the upper edge of the cut in the kth replica to the lower edge of the cut in the lth replica. Repeating this construction for all edges of all cuts (a total of $n\mu$ pastings are required), we finally obtain the desired model of the Riemann surface of $f(z)$.

PROBLEMS

8.1. Let $\{G, f(z)\}$, $\{D, \varphi(z)\}, \ldots$ be the elements of a class $\mathscr{K}$ defining a complete analytic function $F(z)$, in the sense of Definition 4, p. 259. Prove that the class $\mathscr{K}'$ of all elements obtained by differentiating the elements of $\mathscr{K}$, i.e., consisting of all elements of the form $\{G, f'(z)\}$, $\{D, \varphi'(z)\}, \ldots$, also defines a complete analytic function. Discuss the case of higher derivatives.

[49] This follows from the monodromy theorem and its corollary.

Comment. The complete analytic function defined by $\mathscr{K}'$ is called the *derivative* of $F(z)$, and is denoted by $F'(z)$.

8.2. Is there an analytic function $f(z)$ which coincides with the infinitely differentiable real function

$$f(x) = \begin{cases} e^{-1/x^2} & \text{for } x \neq 0, \\ 0 & \text{for } x = 0 \end{cases}$$

on the real axis?

Hint. Recall Problem I.1.6.

8.3. Give an explicit construction of the sequence $\{z_n\}$ figuring in Example 2, p. 261.

Hint. First cover each of the Jordan curves forming the boundary of G by a finite number of disks.

8.4. Show that the domain g'' figuring in Example 2, p. 261 need not be unique.

Hint. Suppose $\zeta \in L$, where the Jordan curve L has distinct end points.

8.5. Given any element $\{G, f\}$, in general noncircular, prove that there is a countable family of circular elements $\{K_n, \varphi_n\} \subset \{G, f\}$ such that

$$\bigcup_{n=1}^{\infty} K_n = G.$$

Hint. Approximate G by an increasing sequence of compact sets $\{F_n\}$, and then cover each set F_1, $F_2 - F_1, \ldots, F_{n+1} - F_n, \ldots$ by a finite number of circular elements.

8.6. Generalize the lemma on p. 267 to the case where the element e is noncircular. Use the result to show that if two circular elements are continuations of each other, they can be joined by a chain consisting entirely of circular elements.

8.7. Verify the existence of the closed Jordan curve figuring in the proof of Theorem 8.7.

Hint. Approximate S_{e_0} by the increasing sequence of open sets $\{\mathcal{O}_n\}$ constructed on p. 81, and choose L to be the boundary of a suitable component of $\mathcal{O}_n$ for sufficiently large n.

8.8. Given a circular element e_0 with center z_0 and star S_{e_0}, let $f(z)$ be the function defined by continuing e_0 along every segment $\overline{z_0 z} \subset S_{e_0}$, just as in Theorem 8.7. Prove that there is a sequence of real numbers $\{\theta_k\}$ and two sequences of positive integers $\{\mu_k\}$ and $\{\nu_k\}$, all independent of $f(z)$, such that

$$\begin{aligned} f(z) &= \lim_{n \to \infty} f_n(z) \\ &= \lim_{n \to \infty} \left[\sum_{k=0}^{\mu_n} \theta_k \frac{f^{(k)}(z_0)}{k!} (z - z_0)^k + \sum_{k=0}^{\nu_n} (1 - \theta_k) \frac{f^{(k)}(z_0)}{k!} (z - z_0)^k \right], \end{aligned}$$

where the sequence $\{f_n(z)\}$ converges uniformly inside S_{e_0}.

8.9. Find the singular points of the following functions:

a) $(\sqrt[6]{z})^4$; b) $(\sqrt[m]{z})^n$ (m, n positive integers); c) $\exp\left(\frac{1}{\sqrt{z}-1}\right)$;

d) $\frac{\sin\sqrt{z}}{\sqrt{z}}$; e) $\tan\frac{1-\sqrt{z}}{1+\sqrt{z}}$; f) $e^{(\operatorname{Ln} z)^n}$ (n a positive integer);

g) $z^i = e^{i \operatorname{Ln} z}$; h) $\sin \operatorname{Ln} z$; i) $\frac{1}{z}\operatorname{Ln}\frac{1}{1-z}$; j) $\frac{1}{z}\operatorname{Arc}\sin z$;

k) $\frac{1}{z^2} + \operatorname{Arc}\tan z$.

Ans.

a) Algebraic branch points of order 2 at $z = 0, \infty$;
c) Algebraic branch points of order 1 at $z = 0, \infty$; one of the branches has an essential singular point at $z = 1$;
e) Algebraic branch points of order 1 at $z = 0, \infty$; one of the branches has a limit point of poles at $z = 1$;
g) Logarithmic branch points at $z = 0, \infty$;
i) Logarithmic branch points at $z = 1, \infty$; every branch except one has a pole of order 1 at $z = 0$;
k) Logarithmic branch points at $z = \pm i$; every branch has a pole of order 2 at $z = 0$.

8.10. Show that the complete analytic function $F(z)$ generated by the element

$$e: \varphi(z) = \sum_{k=1}^{\infty} \frac{z^k}{k} \qquad (|z| < 1)$$

is multiple-valued, without identifying $F(z) = -\operatorname{Ln}(1 - z)$.

Hint. Starting from $z = 0$, continue e along the closed polygonal curve $L = \overline{z_0 z_1} + \cdots + \overline{z_{n-1} z_n}$, where

$$z_j = 1 - e^{-2j\pi i/n} \qquad (j = 0, 1, \ldots, n > 6)$$

(L is a regular n-gon inscribed in the circle $|z - 1| = 1$). Show that one circuit around $z = 1$ in the negative direction causes $F(z)$ to increase by

$$\varphi(z_1) + \varphi\left(\frac{z_2 - z_1}{1 - z_1}\right) + \cdots + \varphi\left(\frac{z_n - z_{n-1}}{1 - z_{n-1}}\right) = 2\pi i.$$

8.11. Let $f(z)$, e_0, ζ and $\overset{\frown}{z_0\zeta}$ be the same as in Theorem 8.9, Corollary 1. Show that if ζ is a regular (singular) point relative to continuation of e_0 along any arc $\overset{\frown}{z_0\zeta}$, then ζ is regular (singular) relative to continuation of e_0 along any other such arc. Prove that if ζ is singular, then it is a unique singular point of $f(z)$ in the sense of Definition 2, p. 283.

8.12. Analyze Problem 8.10 from the standpoint of Example 1, p. 282.

8.13. Give an example where the same point represents both a regular point and a singular point of a complete analytic function.

Hint. Recall Example 2, p. I.393.

8.14. Let G be a bounded simply connected domain, and let e_0 be a circular element with disk $K \subset G$ centered at z_0. Suppose e_0 can be continued along any curve in G, thereby generating a single-valued analytic function $f(z)$ defined on G. Let ζ be an accessible boundary point of G, determined by a Jordan half-interval $\gamma \subset G$ with initial point z_0, and let $\gamma' \subset G$ be any other Jordan half-interval with initial point z_0 determining the same point ζ. Suppose e_0 can be continued along the curve $l = \gamma \cup \{\zeta\}$. Prove that e_0 can also be continued along $l' = \gamma' \cup \{\zeta\}$, with the same result as obtained by continuing e_0 along l.

Hint. First consider the case where γ and γ' have points of intersection in every neighborhood $\mathcal{N}(\zeta)$. More generally, invoke the second condition figuring in the definition on p. 52.

8.15. It follows from the preceding problem that if ζ is a singular point relative to continuation of an element of $f(z)$ along a curve determining ζ, then ζ is also a singular point relative to continuation of an element of $f(z)$ along any other curve determining ζ. Prove that ζ is a *unique* singular point, in the sense of Definition 2, p. 283.

Comment. According to Problems 8.14 and 8.15, every accessible boundary point ζ of G is a regular point of $f(z)$ or a (unique) singular point of $f(z)$, depending on whether or not the elements of $f(z)$ can be continued along curves determining ζ.

8.16. Let the domain G of Problem 8.14 be a disk. Prove that $f(z)$ has a regular (singular) point at ζ in the sense of the comment to Problem 8.15 if and only if the circular element $\{G, f(z)\}$ has a regular (singular) point at ζ in the sense of the definition on p. I.382.

8.17. Paraphrase the results of Sec. I.84 in the language of analytic continuation. In particular, show that a function $f(z)$ defined as the sum of a power series with circle of convergence Γ will have Γ as its natural boundary if and only if the singular points of $f(z)$, as defined in Sec. I.84, are everywhere dense in Γ.

8.18. Let G and $f(z)$ be the same as in Problem 8.14, and suppose G has a prime end, not of the first kind, represented by a continuum F, as in Definition 3, p. 63. Let γ be a Jordan arc with end points on the boundary of G, but with all its other points in G itself. Then γ divides G into two domains (as on p. 59), where the boundary of one of the domains contains F (why?). Denoting this domain by g, suppose every common accessible boundary point of g and G is a singular point of $f(z)$, in the sense of Problem 8.15. Let e_0 be an element of $f(z)$, with disk $K \subset g$ centered at z_0. Show that e_0 cannot be continued along any arc $\overset{\frown}{z_0\zeta}$ joining z_0 to an inaccessible point $\zeta \in F$ if $\overset{\frown}{z\zeta}$ is forbidden to intersect γ.[50]

---

[50] Here, of course, "inaccessible" means "inaccessible from inside G," and there is nothing to prevent us from joining z_0 to ζ by an arc leaving G.

Hint. Since $\widehat{z_0\zeta}$ cannot intersect γ, it must intersect a singular boundary point on its way to ζ.

Comment. Thus, although ζ itself is not a "direct obstacle" to continuation of e_0, obstacles appear along every permissible path of continuation.

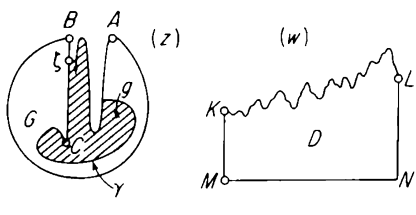


FIGURE 8.4

8.19. Construct an explicit example illustrating the preceding problem.

Ans. Let G be the domain in the z-plane shown in Figure 8.4(a), with boundary consisting of the curve

$$l: y = \sin\frac{1}{x} \qquad (0 < x \leqslant 1)$$

(with no initial point), the segment $x = 0, -1 \leqslant y \leqslant 1$ of the imaginary axis and a circular arc AB joining the points (1, sin 1) and (0, 1). Let D be the domain in the w-plane shown in Figure 8.4(b), with boundary consisting of a Jordan curve KL, none of whose arcs is analytic[51] (hence the indicated "jaggedness") and the polygonal curve $KMNL$. Suppose $w = f(z)$ maps G conformally onto D in such a way that the points A, B (where B is regarded as accessible) and C go into the points L, M and K, respectively (justify the existence of such a mapping). Then the image of the segment BC (regarded as consisting only of accessible points) is the segment MK, and the image of the curve l is the curve KL. Clearly, every point of l is an accessible singular point of $f(z)$. In fact, every arc of l, with equation

$$y = \sin\frac{1}{x} \qquad (0 < \alpha \leqslant x \leqslant \beta \leqslant 1),$$

is a regular analytic curve, in the sense of Sec. 36, and if some point of l is a regular point of $f(z)$, there is an arc of l on which $f(z)$ is analytic and $f'(z)$ is nonvanishing (why?). But then, according to Theorem 7.19, the image of this arc would also be a regular analytic curve, contrary to the construction of KL. Thus every point of l is singular, as asserted.

Now consider the prime end of G, represented by the line segment BC. The only accessible point of BC is the point C, and BC is contained in the boundary of the shaded domain $g \subset G$ shown in Figure 8.4(a). Every other common boundary point of g and G is a point of the curve l, and hence an accessible singular point of $f(z)$. Thus the domains G and g, the function $f(z)$ and the continuum BC (called F in Problem 8.18) have all the required properties.

8.20. Find the branched elements of the function Arc sin z.

Hint. Use the formula

$$\text{Arc sin } z = \int_0^z \frac{dt}{\sqrt{1 - t^2}}.$$

[51] For example, if K and L have abscissas a and b, let KL be the curve with equation $v = \chi(u)$, $a \leqslant u \leqslant b$, where $\chi(u)$ is Weierstrass' continuous, nowhere differentiable function (see p. I.328).

8.21. Given an analytic configuration A, prove that the branch points of A regarded as a Riemann surface (as in Theorem 8.18) coincide with the branched elements of A and have the same orders.

8.22. A single-valued function $f(e)$ defined on a Riemann surface E is said to be *meromorphic* on E if its only singular points on E are poles. Prove that the interior mapping $f(e)$ figuring in Theorem 8.19 is meromorphic on the analytic configuration A.

8.23. Given an analytic configuration A, let $\psi(e)$ be the function which assigns to every element $e_0 \in A$ the value of e_0 at its center. In other words, if e_0 has the series

$$\sum_{n=m}^{\infty} a_n(z - z_0)^{n/k} \quad \text{or} \quad \sum_{n=m}^{\infty} a_n z^{-n/k},$$

then

$$\psi(e_0) = \begin{cases} a_0 & \text{if } m \geqslant 0, \\ \infty & \text{if } m < 0. \end{cases}$$

Prove that $\psi(e)$ is meromorphic on A.

Hint. To verify the analyticity of $\psi(e)$, expand $\psi(e)$ in a neighborhood of any point $e_0 \in A$ as a power series in the local parameter

$$t = \begin{cases} z - z_0 & \text{if } e_0 \text{ is unbranched and } z_0 \text{ is finite,} \\ (z - z_0)^{1/k} & \text{if } e_0 \text{ is branched and } z_0 \text{ is finite,} \\ z^{-1} & \text{if } e_0 \text{ is unbranched and has center } \infty, \\ z^{-1/k} & \text{if } e_0 \text{ is branched and has center } \infty. \end{cases}$$

8.24. Let $w = g(z) \not\equiv \text{const}$ be a function meromorphic in the finite plane E. Then, according to Theorem 7.21, E becomes a Riemann surface when equipped with $w = g(z)$ as an interior mapping. Let A be the analytic configuration consisting of all elements of the inverse function $z = g^{-1}(w)$, so that A becomes a Riemann surface when equipped with the mapping $w = f(e)$ carrying every element $e_0 \in A$ into its center, as in Theorem 8.18. Show that E and A are equivalent Riemann surfaces. Discuss the examples in Sec. 38 from this standpoint.

Hint. The elements of A have series of the form

$$\sum_{n=m}^{\infty} a_n(w - w_0)^{n/k} \quad \text{or} \quad \sum_{n=m}^{\infty} a_n w^{-n/k}.$$

Let $z = \Phi(e)$ be the function carrying every element $e_0 \in A$ into the value of e_0 at its center w_0, i.e., into the value of the function $z = g^{-1}(w)$ at w_0. Then $\Phi(e)$ is a homeomorphism of A onto E carrying each of the interior mappings $f(e)$ and $g(z)$ into the other in the sense of Definition 3, p. 238.

8.25. Prove that $N > 1$ distinct points in the extended plane can be connected by a polygonal curve with $N - 1$ segments and no self-intersections.

8.26. Let $w = f(z)$ be an algebraic function, with Riemann surface E, and let L be a closed curve in the z-plane, with image L' on E. Under what circumstances is L' also closed?

8.27. In Sec. 45 it was shown that an algebraic function is finite-valued and has no singular points except poles and algebraic branch points in the extended plane. Prove that these properties characterize algebraic functions. More precisely, prove the following theorem: Let G be the domain obtained by deleting the points $z_1, \ldots, z_\mu, \infty$ from the extended z-plane, and let $f(z)$ be a function which can be continued analytically along every curve $L \subset G$. Suppose $f(z)$ is at most n-valued and can be represented by a series of the form

$$\sum_{l=m}^{\infty} a_l(z - z_j)^{l/k} \qquad (1 \leqslant k \leqslant n,\ m \leqslant 0)$$

in a neighborhood of each z_j $(j = 1, \ldots, \mu)$ and by a series of the form

$$\sum_{l=m}^{\infty} a_l z^{-l/k} \qquad (1 \leqslant k \leqslant n,\ m \leqslant 0)$$

in a neighborhood of infinity. Then $f(z)$ is an algebraic function.

Hint. Recall the proof of Theorem 8.22.

8.28. Construct the Riemann surfaces of the algebraic function $w = f(z)$ defined by the equation

a) $w^3 - z - 1 = 0$; b) $w^3 - 3w - z = 0$; c) $w^3 + 3w - z = 0$.

9

THE SYMMETRY PRINCIPLE AND ITS APPLICATIONS

46. The Symmetry Principle

The following special technique of analytic continuation is of great importance in the applications:

THEOREM 9.1 (*Symmetry principle*).[1] *Let G be a domain whose boundary contains an accessible open circular arc or line segment γ, and let $g \subset G$ be a domain adjacent to γ.*[2] *Let $w = f(z)$ be meromorphic on G and wide-sense continuous on $g \cup \gamma$, and let G^* be the domain symmetric to G with respect to γ.*[3] *Moreover, suppose $\Gamma = f(\gamma)$ is itself a circular arc or line segment, and let $f^*(z)$ be the function defined on $G^* \cup \gamma$ by the conditions*

1. *$f^*(z) = f(z)$ for all $z \in \gamma$;*
2. *$f^*(z^*)$ and $f(z)$ are symmetric with respect to Γ for all $z^* \in G^*$, where z^* is the point symmetric to $z \in G$ with respect to γ.*

Then $f^*(z)$ is the analytic continuation of $f(z)$ from g (or G) across γ into G^* in the sense that the function $\varphi(z)$ which equals $f(z)$ on G and $f^*(z)$ on G^* is meromorphic on the domain $G \cup \gamma \cup G^*$.[4]

---

[1] Often called *Schwarz's reflection principle.*

[2] By an *open* arc, we mean an arc minus its end points. The concept of a domain adjacent to γ has the same meaning as on p. 71. Thus g is bounded by a closed Jordan curve, say $\gamma \cup l$.

[3] More exactly, symmetric (in the sense of Sec. I.48) with respect to the circle or straight line containing γ.

[4] We might call $f^*(z)$ a "wide-sense analytic continuation" of $f(z)$.

Proof. First we make a number of preliminary observations:

1. The domain g may coincide with G as in Figure 9.1(a), or g may be a proper subset of G as in Figure 9.1(b).
2. In asserting that $G \cup \gamma \cup G^*$ is a domain, the accessibility of γ is crucial (why?).
3. If $G \cap G^* \neq 0$ (as in Figure 9.1), then the function equal to $f(z)$ on G and to $f^*(z)$ on G^* will be double-valued on $G \cap G^*$, unless $f(z) \equiv \text{const}$, and in fact the two values will be symmetric with respect to Γ.
4. The fact that $f(z)$ and $f^*(z)$ are allowed to be meromorphic, instead of just analytic, is a natural feature of the problem. In fact, let Γ be a circular arc and suppose that $f(z)$, without necessarily having poles in G, maps a point $z_0 \in G$ into the center of the circle containing Γ. Then $f^*(z)$ becomes infinite at the point z_0^* symmetric to z_0 with respect to γ, and hence has a pole at z_0^*.[5]

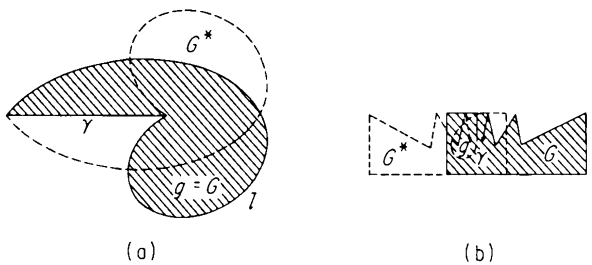


FIGURE 9.1

Turning to the proof itself, we note that there is no loss of generality in assuming that γ and Γ are both segments of the real axis. In fact, this configuration can always be achieved by making suitable preliminary Möbius transformations in both the z and w-planes, and such transformations (and their inverses) preserve wide-sense continuity, symmetry and accessibility, and moreover carry meromorphic functions into meromorphic functions. Thus we can write

$$z = x, \qquad y = \operatorname{Im} z = 0$$

on γ and

$$w = f(z) = u, \qquad v = \operatorname{Im} f(z) = 0$$

on Γ. Then the points z^* and z symmetric with respect to γ are just complex conjugates of each other, and the same is true of the points w and w^* symmetric with respect to Γ:

$$z^* = \bar{z}, \qquad w^* = \bar{w}.$$

[5] Here we anticipate the fact that z_0^* is an isolated singular point of $f^*(z)$.

Given any point $z_0 \in G$, let ρ be the distance between z_0 and the nearest pole $\zeta \neq z_0$ of $f(z)$ in G.[6] Being meromorphic on G, $f(z)$ has an expansion at z_0 of the form

$$f(z) = \sum_{n=m}^{\infty} a_n(z - z_0)^n \qquad (0 < |z - z_0| < \rho).$$

But then

$$f^*(z^*) = \overline{f(z)} = \sum_{n=m}^{\infty} \bar{a}_n(\bar{z} - \bar{z}_0)^n = \sum_{n=m}^{\infty} \bar{a}_n(z^* - z_0^*)^n \qquad (0 < |z^* - z_0^*| < \rho).$$

Therefore f^* is meromorphic on G^*,[7] since the point $z_0 \in G$ (and hence the point $z_0^* \in G^*$) is arbitrary.

Next we show that f^* is (wide-sense) continuous on $g^* \cup \delta$, where g^* is the domain symmetric to g with respect to γ (f^* is obviously continuous on $g^* \subset G^*$). Let $z = x_0$ be any point of γ. Then

$$\lim_{\substack{z \to x_0 \\ z \in g}} f(z) = f(x_0) = u(x_0, 0),$$

since $f(z) = u(x, y) + iv(x, y)$ is real on γ and continuous on $g \cup \gamma$, and hence

$$\lim_{\substack{(x,y) \to (x_0,0) \\ (x,y) \in g}} v(x, y) = 0.$$

It follows that

$$\lim_{\substack{z^* \to x_0 \\ z^* \in g}} f^*(z^*) = \lim_{\substack{z \to x_0 \\ z \in g}} \overline{f(z)} = \lim_{\substack{(x,y) \to (x_0,0) \\ (x,y) \in g}} [u(x, y) - iv(x, y)] = u(x_0, 0) = f(x_0) = f^*(x_0),$$

i.e., f^* is continuous on $g^* \cup \gamma$, as asserted (how about the case where z^* approaches x_0 along γ itself?).

Finally we show that the function $\varphi(z)$ which equals $f(z)$ on $G \cup \gamma$ and $f^*(z)$ on $G^* \cup \gamma$ is meromorphic on γ. Let $z_0 = x_0$ be any point of γ such that $\varphi(x_0) \neq \infty$.[8] Since x_0 is accessible, there is a disk K with center x_0 such that $K \subset g \cup \gamma \cup g^*$ (why?). Let K_1 and K_2 be the intersections of K with the open upper and lower half-planes, respectively. Replacing K by a smaller disk if necessary, we can assume that $\varphi(z)$ is single-valued and analytic on K_j and continuous on $\overline{K}_j$ for $j = 1, 2$. Let Δ be any closed

---

[6] If there is no such pole, let ρ be the distance between z_0 and the boundary of G. Note that z_0 itself may be a pole (if $m < 0$).

[7] This is a situation where confusion can be avoided by omitting the argument of the function f^*.

[8] If $x_0 = \infty$, we make the preliminary transformation $\zeta = 1/z$ carrying x_0 into the origin.

triangle contained in K. Then there are three possibilities:

1. Δ is contained in either K_1 or K_2;
2. Δ is contained in either $\overline{K}_1$ or $\overline{K}_2$;
3. The real axis cuts Δ into two closed convex polygons Δ_1 and Δ_2, one contained in $\overline{K}_1$ and the other in $\overline{K}_2$.

In every case, we have

$$\int_L \varphi(z)\, dz = 0,$$

where L is the boundary of Δ, traversed in the positive direction. In the first case, this follows from Cauchy's integral theorem, in the second case from Theorem I.13.3 (or the stronger Theorem 3.10),[9] and in the third case from Theorem I.13.3 and the observation that the two integrals along the common side of Δ_1 and Δ_2 cancel each other out. But then, by Morera's theorem (in particular, the remark on p. I.302), $\varphi(z)$ is analytic on K, i.e., in a neighborhood of x_0. Since x_0 is an arbitrary point of γ at which $\varphi(x_0) \neq \infty$, it follows that $\varphi(z)$ is meromorphic on $G \cup \gamma \cup G^*$, and the theorem is proved.

Next we prove a number of important implications of the symmetry principle.

THEOREM 9.2. *Let G be a domain whose boundary contains an accessible open regular analytic arc $\gamma: z = \lambda(t)$, $a < t < b$, and let $g \subset G$ be a domain adjacent to γ. Let $w = f(z)$ be meromorphic on G and wide-sense continuous on $g \cup \gamma$, and suppose $\Gamma = f(\gamma)$ is itself a regular analytic arc with equation $w = \Lambda(\tau)$, $\alpha < \tau < \beta$. Then $f(z)$ can be continued analytically from g across γ.*

Proof. If $t_0 \in (a, b)$ and $z_0 = \lambda(t_0)$, then $t = \lambda^{-1}(z)$ maps a neighborhood $\mathcal{N}(z_0)$ conformally onto a domain d containing t_0, while carrying the arc $\gamma \cap \mathcal{N}(z_0)$ containing z_0 into a segment δ of the real axis containing t_0. Similarly, if $\tau_0 \in (\alpha, \beta)$ and $w_0 = \Lambda(\tau_0)$, then $\tau = \Lambda^{-1}(w)$ maps a neighborhood $\mathcal{N}(w_0)$ conformally onto a domain D containing τ_0, while carrying the arc $\Gamma \cap \mathcal{N}(w_0)$ containing w_0 into a segment of the real axis containing τ_0. Thus the function $\tau = \Lambda^{-1} f \lambda(t)$ is meromorphic on some domain d', whose boundary contains the accessible line segment δ. Moreover $\Lambda^{-1} f \lambda(t)$ is continuous on $d' \cup \delta$, and maps δ into a segment of the real axis. It follows from the symmetry principle that $\Lambda^{-1} f \lambda(t)$ can be continued analytically from d' across δ. Therefore the function

$$f(z) = \Lambda(\Lambda^{-1} f \lambda)\lambda^{-1}(z)$$

can be continued analytically from g across γ, as asserted.

---

[9] Theorem I.13.3 and 3.10 are generalizations of Cauchy's integral theorem.

Remark. Since circular arcs and line segments are special cases of regular analytic arcs (see Example 1, p. 242), Theorem 9.2 is a generalization of Theorem 9.1.

THEOREM 9.3. *Under the hypotheses of Theorem* 9.2, *suppose* $f(z)$ *vanishes on* γ. *Then* $f(z)$ *vanishes identically.*

Proof. According to Theorem 9.2, $f(z)$ is analytic on a larger domain $G' \supset G$ containing γ. Now apply the interior uniqueness theorem.

Remark. The example on p. II.213 is a special case of Theorem 9.3, which in turn is a special case of the Luzin-Privalov boundary uniqueness theorem cited on p. I.371.

THEOREM 9.4. *Let* G *be a simply connected domain whose boundary contains an accessible regular analytic arc* γ, *and let* $w = f(z)$ *be a function mapping* G *conformally onto a disk or a half-plane. Then* $f(z)$ *can be continued analytically across* γ *from any domain* $g \subset G$ *adjacent to* γ.

Proof. According to Theorem 2.25, $f(z)$ is continuous on $\bar{g}$ as well as analytic on g. Now apply Theorem 9.3.

THEOREM 9.5. *Let* G *and* D *be two Jordan domains, with boundaries* Γ *and* Δ, *respectively. Then there exists a unique function* $w = f(z)$ *which maps* G *conformally onto* D *and carries three given points* $z_1, z_2, z_3 \in \Gamma$ *into three given points* $w_1, w_2, w_3 \in \Delta$, *provided the curves* $\overset{\frown}{z_1z_2} + \overset{\frown}{z_2z_3} + \overset{\frown}{z_3z_1}$ *and* $\overset{\frown}{w_1w_2} + \overset{\frown}{w_2w_3} + \overset{\frown}{w_3w_1}$ *have the same orientation.*

Proof. By the corollary to Theorem 2.24, $f(z)$ is continuous on $\bar{G}$ and establishes a homeomorphism between Γ and Δ. Therefore, just as on p. II.119, $w = f(z)$ traverses Δ in the same direction as z traverses Γ (here we use the argument principle generalized in the same way as Theorem 3.10 generalizes Cauchy's integral theorem). This explains the stipulation about preservation of orientation.

We begin by considering the case where G and D are the unit disks $|z| < 1$ and $|w| < 1$. Suppose that among the infinitely many functions mapping G conformally onto D, there are two functions $f(z)$ and $g(z)$ such that

$$f(z_j) = w_j, \quad g(z_j) = w_j \qquad (j = 1, 2, 3).$$

By the symmetry principle, the functions $f(z)$ and $g(z)$ can be continued across the unit circle $\gamma: |z| = 1$ into the domain symmetric to $|z| < 1$ with respect to γ, i.e., the domain $|z| > 1$. Therefore $f(z)$ and $g(z)$ are meromorphic and in fact *univalent* (why?) in the whole finite plane. It follows from Theorem II.2.4 that $f(z)$ and $g(z)$ are Möbius transformations. But a Möbius transformation is uniquely specified by its values at three distinct points (recall Theorem II.10.6), and hence $f(z) \equiv g(z)$.

Returning to the case of arbitrary Jordan domains G and D, let $\zeta = \varphi(z)$ and $\omega = \psi(w)$ be two functions mapping G and D conformally onto the unit disks $|\zeta| < 1$ and $|\omega| < 1$, and let

$$\varphi(z_j) = \zeta_j, \quad \psi(w_j) = \omega_j \qquad (j = 1, 2, 3).$$

Moreover, let $\omega = F(\zeta)$ be the unique conformal mapping of $|\zeta| < 1$ onto $|\omega| < 1$ such that

$$F(\zeta_j) = \omega_j \qquad (j = 1, 2, 3)$$

[the existence and uniqueness of $F(\zeta)$ were just proved]. Then obviously

$$w = f(z) = \psi^{-1}F\varphi(z)$$

maps G conformally onto D and carries z_1, z_2, z_3 into w_1, w_2, w_3. If there were another conformal mapping of G onto D carrying z_1, z_2, z_3 into w_1, w_2, w_3, then there would be another function

$$\omega = G(\zeta) = \psi g\varphi^{-1}(\zeta)$$

mapping $|\zeta| < 1$ conformally onto $|\omega| < 1$ and carrying $\zeta_1, \zeta_2, \zeta_3$ into $\omega_1, \omega_2, \omega_3$. Since this is impossible, $f(z) \equiv g(z)$ and the theorem is proved.

COROLLARY. *Theorem* 9.5 *remains valid if* Γ *is a closed unbounded Jordan curve*[10] *and* G *is either of the domains with boundary* Γ, *and also if* Δ *is a closed unbounded Jordan curve and* D *is either of the domains with boundary* Δ.

Proof. Make preliminary Möbius transformations carrying G and D into Jordan domains.

47. More on the Schwarz-Christoffel Transformation

In Sec. II.20 we used the Schwarz-Christoffel transformation to map the upper half-plane onto a bounded polygon, without, however, actually deriving the basic formula (II.4.14). We now have enough theory at our disposal to deduce the formula from general principles of conformal mapping:

THEOREM 9.6 (*Schwarz-Christoffel transformation for bounded polygons*). *Suppose* $w = f(z)$ *maps the upper half-plane* $\Pi_U: \operatorname{Im} z > 0$ *conformally onto the interior of a (bounded) closed polygonal Jordan curve* Λ *with interior angles* $\alpha_1\pi, \ldots, \alpha_n\pi$ $(0 < \alpha_j < 2, j = 1, \ldots, n)$, *and suppose*

[10] I.e., a curve whose image on the Riemann sphere passes through the north pole (cf. pp. I.333 and II.120).

the points $a_1, \ldots, a_n$ $(-\infty < a_1 < \cdots < a_n < +\infty)$ *correspond to the vertices of* Λ. *Then*

$$w = f(z) = C \int_{z_0}^{z} \prod_{j=1}^{n} (t - a_j)^{\alpha_j - 1}\, dt + C_1, \tag{9.1}$$

where z_0 *is any point of* $\overline{\Pi}_U$ *and* C, C_1 *are complex constants.*[11]

Proof. By the corollary to Theorem 2.24, $f(z)$ is continuous on the closed upper half-plane $\overline{\Pi}_U$ and establishes a homeomorphism between the real axis and Λ. Let $w_1, \ldots, w_n$ be the vertices of Λ (the images of $a_1, \ldots, a_n$), arranged in order of their occurrence as Λ is traversed in the positive direction. We can always choose the points $a_1, \ldots, a_n$ to be finite, since otherwise the transformation

$$z' = \frac{1}{a - z},$$

where a is a point of the real axis distinct from $a_1, \ldots, a_n$, transforms Π_U into itself and $f(z)$ into a new mapping under which the inverse images of the vertices $w_1, \ldots, w_n$ are finite. Using the symmetry principle, applied to the function $w = f(z)$ and the intervals $(a_1, a_2) \ldots, (a_n, a_1)$ of the real axis (one of which contains the point at infinity), we find that every point of the real axis is a regular point of $f(z)$, except possibly the points $a_1, \ldots, a_n$ themselves.

Next we draw a circle γ_j with center w_j and radius so small that $I(\gamma_j)$ contains no vertices of Λ other than w_j. Consider the circular sector $g_j = I(\gamma_j) \cap I(\Lambda)$, whose vertex angle equals $\alpha_j \pi$ radians. Any single-valued branch of the function

$$\tau = (w - w_j)^{1/\alpha_j}$$

maps g_j onto a half-disk K_j with its center at the origin. Correspondingly, there is a domain $\mathcal{G}_j \subset \Pi_U$, whose boundary contains an accessible segment δ of the real axis, which is mapped conformally onto K_j by the function

$$\tau = [f(z) - w_j]^{1/\alpha_j} \equiv f_j(z).$$

This mapping is one-to-one and continuous on $\overline{\mathcal{G}}_j$, and carries δ into the diameter of K_j. Applying the symmetry principle to the domain $\mathcal{G}_j$ and the function $f_j(z)$, we find that $\tau = f_j(z)$ is analytic on a domain containing the point a_j, which it maps conformally onto a neighborhood of the point $\tau = 0$. Therefore $f_j(z)$ has an expansion at a_j of the form

$$f_j(z) = A_j(z - a_j) + \cdots \qquad (A_j \neq 0),$$

---

[11] Note that apart from a trivial translation and rotation, formulas (9.1) and (II.4.14) are identical (there is no loss of generality in setting $z_0 = 0$).

which implies

$$f(z) = w_j + [f_j(z)]^{\alpha_j} = w_j + (z - a_j)^{\alpha_j}[B_j + C_j(z - a_j) + \cdots] \qquad (B_j \neq 0).$$

It follows that a_j is a branch point of $f(z)$ of finite order if α_j is rational and of infinite order if α_j is irrational.

The fact that $f(z)$ has no singular points other than those just found will now be used to find an integral representation of $f(z)$. Differentiating $f(z)$ twice with respect to z, we obtain

$$f'(z) = B_j\alpha_j(z - a_j)^{\alpha_j - 1} + C_j(\alpha_j + 1)(w - a_j)^{\alpha_j} + \cdots,$$
$$f''(z) = B_j\alpha_j(\alpha_j - 1)(z - a_j)^{\alpha_j - 2} + C_j(\alpha_j + 1)\alpha_j(z - a_j)^{\alpha_j - 1} + \cdots,$$

and hence

$$\frac{f''(z)}{f'(z)} = \frac{\alpha_j - 1}{z - a_j} + \cdots,$$

i.e., the function $f''(z)/f'(z)$ has simple poles at the points a_j with residues equal to $\alpha_j - 1$ $(j = 1, \ldots, n)$. Moreover, let a_0 be any finite point distinct from $a_1, \ldots, a_n$. Then $f(z)$ is analytic at a_0 and moreover $f'(a_0) \neq 0$, since $f(z)$ is one-to-one in a neighborhood of a_0. Therefore $f''(z)/f'(z)$ has no finite singular points other than the poles at $a_1, \ldots, a_n$. Since $f(z)$ is analytic at infinity and one-to-one in a neighborhood of infinity (why?), we have

$$f(z) = D_0 + \frac{D_1}{z} + \frac{D_2}{z^2} + \cdots \qquad (D_1 \neq 0),$$

$$f'(z) = -\frac{D_1}{z^2} + \cdots, \qquad f''(z) = \frac{2D_1}{z^3} + \cdots,$$

and hence

$$\frac{f''(z)}{f'(z)} = -\frac{2}{z} + \cdots, \tag{9.2}$$

i.e., $z = \infty$ is a regular point of $f''(z)/f'(z)$.

Thus, finally, subtracting the sum

$$\sum_{j=1}^{n} \frac{\alpha_j - 1}{z - a_j}$$

from $f''(z)/f'(z)$, we obtain a function which is single-valued and analytic everywhere in the extended plane (the single-valuedness follows from the monodromy theorem. But then by Liouville's theorem (in particular, the remark on p. I.364),

$$\frac{f''(z)}{f'(z)} - \sum_{j=1}^{n} \frac{\alpha_j - 1}{z - a_j} \equiv \text{const} = c. \tag{9.3}$$

Setting $z = \infty$ in (9.3) and using (9.2), we find that $c = 0$. Therefore integration of (9.3) gives

$$f'(z) = C \prod_{j=1}^{n} (z - a_j)^{\alpha_j - 1},$$

and another integration yields the required representation (9.1).

COROLLARY. *Of the n points $a_1, \ldots, a_n$, three can be chosen arbitrarily, provided they have the same order as the corresponding vertices. The others are then uniquely determined by Λ, in particular by the $n - 2$ independent conditions*

$$|C| \int_{a_{j-1}}^{a_j} |(t - a_1)^{\alpha_1 - 1} \cdots (t - a_n)^{\alpha_n - 1}| \, dt = l_j \qquad (j = 1, \ldots, n - 2),$$

where l_j is the length of the side associated with the angles $\alpha_{j-1}\pi$ and $\alpha_j\pi$ ($\alpha_0 = \alpha_n$).

Proof. Use Theorem 9.5 and its Corollary, recalling Example 3, p. II.134.

Example 1. Suppose Λ is a triangle with vertices w_1, w_2, w_3, so that

$$\Lambda = \overline{w_1 w_2} + \overline{w_2 w_3} + \overline{w_3 w_1}.$$

Then the inverse images a_1, a_2, a_3 can be chosen arbitrarily, provided only that $a_1 < a_2 < a_3$, and (9.1) takes the form

$$w = f(z) = C \int_{z_0}^{z} (t - a_1)^{\alpha_1 - 1}(t - a_2)^{\alpha_2 - 1}(t - a_3)^{\alpha_3 - 1} \, dt + C_1,$$

where we choose any branch of the integrand which is single-valued on Π_U. To determine the constants C and C_1,[12] we set $z = a_1$, $z = a_2$, obtaining the conditions

$$w_j = C \int_{z_0}^{a_j} (t - a_1)^{\alpha_1 - 1}(t - a_2)^{\alpha_2 - 1}(t - a_3)^{\alpha_3 - 1} \, dt + C_1 \qquad (j = 1, 2),$$

which can be solved for C and C_1. The values of $f(z)$ on the real axis can be obtained by analytic continuation, which in this case amounts to using the continuity of $f(z)$ on $\overline{\Pi}_U$.

Example 2. Let Λ be a quadrilateral with vertices w_1, w_2, w_3, w_4, so that

$$\Lambda = \overline{w_1 w_2} + \overline{w_2 w_3} + \overline{w_3 w_4} + \overline{w_4 w_1}.$$

Of the four inverse images a_1, a_2, a_3, a_4 ($a_1 < a_2 < a_3 < a_4$), only three can now be chosen arbitrarily, say a_1, a_2, a_3. Then (9.1) takes the form

$$w = f(z) = C \int_{z_0}^{z} (t - a_1)^{\alpha_1 - 1}(t - a_2)^{\alpha_2 - 1}(t - a_3)^{\alpha_3 - 1}(t - a_4)^{\alpha_4 - 1} \, dt + C_1,$$

[12] The point z_0 is regarded as fixed in advance (cf. footnote 11, p. 321).

where a_4, C and C_1 are determined from the conditions

$$w_j = C\int_{z_0}^{a_j}(t-a_1)^{\alpha_1-1}(t-a_2)^{\alpha_2-1}(t-a_3)^{\alpha_3-1}(t-a_4)^{\alpha_4-1}\,dt + C_1 \qquad (j=1,2,3).$$

Things can be simplified in the case where Λ has an axis of symmetry. For example, let $I(\Lambda)$ be the rectangular domain

$$-A < u < A, \qquad 0 < v < B$$

($w = u + iv$), and suppose $w = f(z)$ is the function mapping the first quadrant of the z-plane into the right-hand half of $I(\Lambda)$, i.e., the domain

$$0 < u < A, \qquad 0 < v < B,$$

in such a way that the points 0, 1, ∞ go into the points 0, A, iB. Then the inverse image of the point $A + iB$ is a point of the positive real axis lying between 1 and ∞. Let this point be denoted by $1/k$, where $0 < k < 1$. Applying the symmetry principle to the first quadrant of the z-plane and the function $w = f(z)$, with the positive imaginary axis playing the role of the accessible line segment, we find that the same function (after analytic continuation) maps the upper half-plane onto the original rectangle $I(\Lambda)$ and carries the points ± 1, $\pm 1/k$ into the vertices $\pm A$, $\pm A \pm iB$. Therefore, since $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \frac{1}{2}$ for a rectangle, (9.4) takes the symmetric form

$$w = f(z) = C\int_{z_0}^{z}(t-1)^{-1/2}\left(t-\frac{1}{k}\right)^{-1/2}\left(t+\frac{1}{k}\right)^{-1/2}(t+1)^{-1/2}\,dt + C$$

$$= C'\int_0^z \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}} + C_1' \qquad (0 < k < 1).$$

Setting $C' = 1$ and $C_1' = 0$, we obtain the elliptic integral of the first kind studied in Sec. II.19 [cf. (II.4.7)].

Next we generalize the Schwarz-Christoffel transformation to the case (of great practical importance) where Λ is a "generalized polygonal curve" in the extended w-plane, whose sides may be traversed more than once and are allowed to have infinite length. Correspondingly, the angles $\alpha_1\pi, \ldots, \alpha_n\pi$ will now be permitted to take the values 0, $\pm 2\pi$. Some examples of such polygonal curves are shown in Figure 9.2 for the cases $n = 1, 2, 3, 4$, where n is the number of sides of Λ. In every case where $z = \infty$ is a point of Λ (as in all the cases shown in Figure 9.2), we agree to regard ∞ as a vertex of the polygon with boundary Λ, assigning ∞ a negative "interior" angle (unless the angle vanishes). For example, the polygon in Figure 9.2(a) has a single angle of $-\pi$ radians with its vertex at infinity, the polygon in Figure 9.2(b) has one angle of 2π radians with a finite vertex and another angle of -2π radians with its vertex at infinity, the polygon in Figure 9.2(c) has two angles of 0 radians with their vertices at infinity, and so on. It can easily be verified

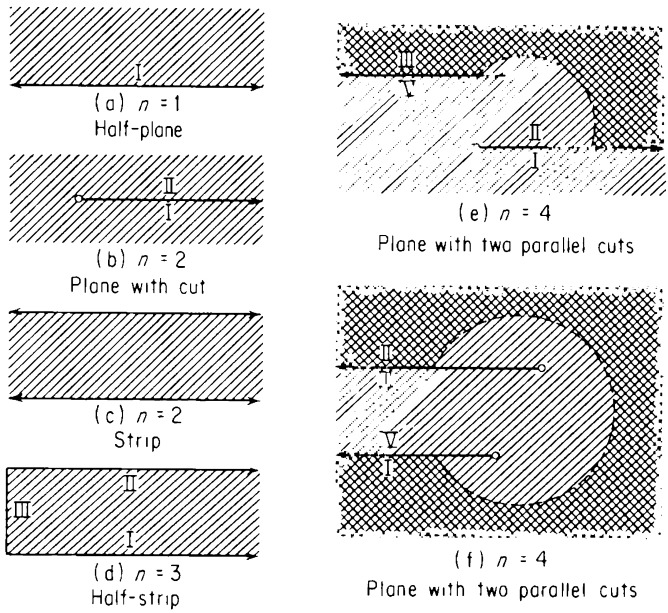


FIGURE 9.2

that the sum of the angles of each polygon shown in Figure 9.2 equals $(n - 2)\pi$, a result which holds whenever the point at infinity is not an interior point of the polygon (see the corollary to Theorem 9.7).

THEOREM 9.7 (*Schwarz-Christoffel transformation for polygons with* ∞ *as a boundary point*). *Suppose* $w = f(z)$ *maps the upper half-plane* Π_U: Im $z > 0$ *conformally onto the domain with boundary* Λ, *where* Λ *is a polygonal curve containing the point at infinity, with angles* $\alpha_1\pi, \ldots,$ $\alpha_n\pi$ *in the interval* $[0, 2\pi]$ *at finite vertices and in the interval* $[-2\pi, 0]$ *at infinite vertices. Then* $w = f(z)$ *is again given by formula* (9.1).[13]

Proof. The proof resembles that of Theorem 9.6 except that now we must consider a number of new possibilities:

1. *There is an angle of* 2π *radians with its vertex at a finite point* w_j. In this case, $\alpha_j = 2$ and the inverse image a_j is a regular point of $w = f(z)$, with an expansion at a_j of the form

$$f(z) = w_j + B_j(z - a_j)^2 + \cdots \qquad (B_j \neq 0).$$

2. *There is an angle of* $\alpha_j\pi$ *radians with its vertex at* ∞ *whose sides are nonparallel lines or rays.* Then $-2 < \alpha_j < 0$ $(\alpha_j \neq -1)$, and the transformation $\tau = w^{1/\alpha_j}$ maps the infinite domain $g_j \subset G$ whose

[13] In particular, the points $a_1, \ldots, a_n, z_0$ and the constants C, C_1 have the same meaning as in Theorem 9.6. Moreover, the analogue of the corollary to Theorem 9.6 holds for Theorems 9.7, 9.9, 9.11 and 9.12.

boundary consists of the sides of the angle and an arc of a circle with center $z = 0$ onto a half-disk with center $\tau = 0$. The fact that a_j is a regular point now follows in the usual way from the symmetry principle.

3. *There is an angle of $\alpha_j\pi$ radians with its vertex at ∞ whose sides lie on the same line.* Then the angle is either $-\pi$ as in Figure 9.2(a) or -2π as in Figure 9.2(b). Here again a_j is a regular point, as we see at once by making the transformation $\tau = 1/z$ or $\tau = 1/z^2$.

4. *There is an angle of $\alpha_j\pi$ radians with its vertex at ∞ whose sides are distinct parallel lines or rays.* Then the angle is either 0 as in Figures 9.2(c), 9.2(d) and 9.2(f), $-\pi$ as in Figure 9.2(e), or -2π as in Figure 9.2(f) again. The difficulty of the proof all centers on this case, which we further divide into three subcases:

   a. *First let $\alpha_j = 0$, as in Figure* 9.2(c). There is no loss of generality in assuming that the parallel sides are segments of the lines $\operatorname{Im} w = \pm h$, $h > 0$ (this can always be achieved by making a preliminary entire linear transformation). Drawing a line segment perpendicular to the parallel sides, we obtain an infinite subdomain $g_j \subset G$, which is mapped by the transformation

$$\tau = e^{cw} \qquad \left(c = -\frac{2\pi}{h}\right)$$

   onto a half-disk with its center at $\tau = 0$ (the image of the point at infinity). Therefore, applying the symmetry principle, we find that

$$\tau = e^{cw} = e^{cf(z)} \equiv \varphi(z)$$

   is analytic in a neighborhood of a_j. It follows that

$$\varphi(z) = A_j(z - a_j) + B_j(z - a_j)^2 + \cdots,$$

   where $A_j \neq 0$ (why?). Therefore

$$\varphi'(z) = ce^{cf(z)}f'(z) = A_j + 2B_j(z - a_j) + \cdots,$$
$$\varphi''(z) = c^2e^{cf(z)}[f'(z)]^2 + ce^{cf(z)}f''(z) = 2B_j + \cdots,$$

   and hence

$$cf'(z) = \frac{A_j + 2B_j(z - a_j) + \cdots}{A_j(z - a_j) + B_j(z - a_j)^2 + \cdots} = \frac{1}{z - a_j} + \cdots,$$
$$cf'(z) + \frac{f''(z)}{f'(z)} = \frac{2B_j + \cdots}{A_j + 2B_j(z - a_j) + \cdots} = C_j + \cdots.$$

   But then

$$\frac{f''(z)}{f'(z)} = (C_j + \cdots) - \frac{1}{z - a_j} + \cdots = -\frac{1}{z - a_j} + \cdots.$$

Thus $f''(z)/f'(z)$ has a simple pole at a_j with residue $-1 = \alpha_j - 1$, just as in the case of an angle with its vertex at a finite point.

b. *Next let* $\alpha_j = -1$, *as in Figure* 9.2(e). Consider the subdomain $g_j \subset G$ bounded by the sides of the angle and a circular arc, i.e., the cross-hatched domain shown in Figure 9.2(e). Making the transformations

$$\eta = cw + d, \quad \eta = \tau' + e^{\tau'}, \quad \tau = e^{-\tau'}$$

(recall the remark on p. I.239), we consecutively map g_j onto each of the domains shown in Figures 9.3(a), 9.3(b) and 9.3(c).

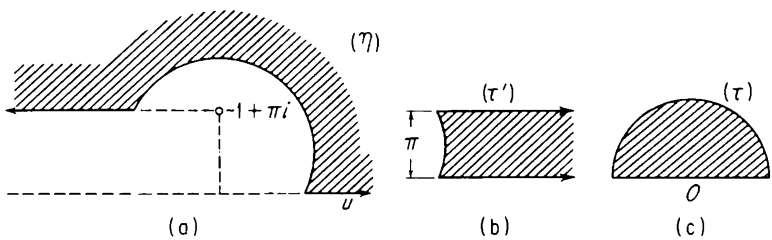


FIGURE 9.3

The result is a function $\tau = \varphi(z)$ with an expansion at a_j of the form

$$\varphi(z) = A_j(z - a_j) + B_j(z - a_j)^2 + \cdots \qquad (A_j \neq 0).$$

Differentiating the formula

$$cf(z) + d = \frac{1}{\varphi(z)} - \operatorname{Ln} \varphi(z)$$

twice with respect to z, we obtain

$$cf'(z) = -\frac{\varphi'(z)}{[\varphi(z)]^2} - \frac{\varphi'(z)}{\varphi(z)} = -\frac{1}{A_j(z - a_j)^2} + \cdots,$$

$$cf''(z) = \frac{2}{A_j(z - a_j)^3} + \cdots,$$

and hence

$$\frac{f''(z)}{f'(z)} = -\frac{2}{z - a_j} + \cdots,$$

i.e., the function $f''(z)/f'(z)$ has a simple pole at a_j with residue $-2 = \alpha_j - 1$.

c. *Finally let* $\alpha_j = -2$, *as in Figure* 9.2(f). Consider the subdomain $g_j \subset G$ bounded by the sides of the angle and a

circular arc, i.e., the cross-hatched domain shown in Figure 9.2(e). Making the transformations

$$\eta = cw + d, \quad \eta = \tau' + e^{\tau'}, \quad \tau = e^{-\tau'/2},$$

we consecutively map g_j onto each of the domains shown in Figures 9.4(a), 9.4(b) and 9.4(c). The result is a function $\tau = \varphi(z)$ with an expansion at a_j of the form

$$\varphi(z) = A_j(z - a_j) + B_j(z - a_j)^2 + \cdots \qquad (A_j \neq 0),$$

satisfying the relation

$$cf(z) + d = \frac{1}{[\varphi(z)]^2} - 2 \operatorname{Ln} \varphi(z).$$

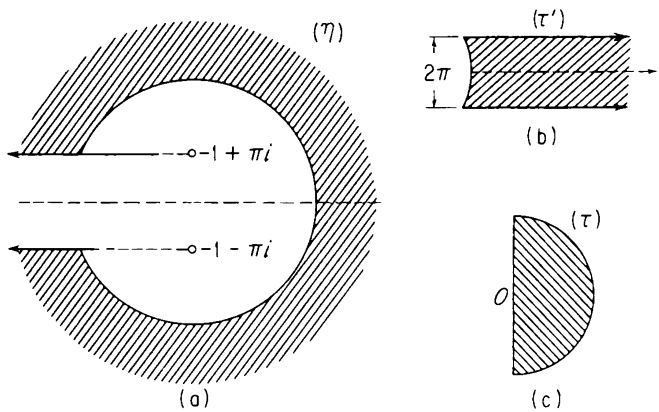


FIGURE 9.4

Then, just as before, we find that

$$\frac{f''(z)}{f'(z)} = -\frac{3}{z - a_j} + \cdots,$$

i.e., the function $f''(z)/f'(z)$ has a simple pole at a_j with residue $-3 = \alpha_j - 1$.

Thus, in every case, the inverse image a_j of the vertex w_j is a simple pole of $f''(z)/f'(z)$, with residue $\alpha_j - 1$. The proof of Theorem 9.6 now follows by exactly the same argument as used to prove Theorem 9.6.

THEOREM 9.8. *Let G be a polygon in the extended plane which does not contain ∞ as an interior point. Then the sum of the vertex angles of G equals $(n - 2)\pi$.*

Proof. In both Theorems 9.6 and 9.7, the function $f''(z)/f'(z)$ is a rational function, whose residue at infinity equals 2 (recall the calcula-

tion on p. 322). But the sum of all the residues of a rational function equals zero (see Theorems II.2.6 and II.10.9), i.e.,

$$\sum_{j=1}^{n} (\alpha_j - 1) + 2 = 0,$$

or

$$\sum_{j=1}^{n} \alpha_j = n - 2.$$

There is still another case to be considered:

THEOREM 9.9 (*Schwarz-Christoffel transformation for polygons with ∞ as an interior point*). *Let $f(z)$, Λ and $\alpha_1, \ldots, \alpha_n$ be the same as in Theorem 9.7, except that now Λ does not contain the point at infinity. Then formula* (9.1) *is replaced by*

$$w = f(z) = C \int_{z_0}^{z} \frac{\prod_{j=1}^{n} (t - a_j)^{\alpha_j - 1}}{(t - \beta)^2 (t - \bar{\beta})^2} \, dt + C_1, \tag{9.4}$$

where the points $a_1, \ldots, a_n$ ($-\infty < a_1 < \cdots < a_n < +\infty$) are the inverse images of the vertices of the polygon and β (Im $\beta > 0$) is the inverse image of the point at infinity.

Proof. As before, $f''(z)/f'(z)$ has a simple pole with residue $\alpha_j - 1$ at every a_j, but now there are two extra poles at β and its complex conjugate $\bar{\beta}$. Since the mapping $f(z)$ is one-to-one, β must be a simple pole, and hence the expansion of $f(z)$ at β is of the form

$$f(z) = \frac{A}{z - \beta} + \cdots,$$

which implies

$$\frac{f''(z)}{f'(z)} = -\frac{2}{z - \beta} + \cdots.$$

Applying the symmetry principle to $f(z)$ and the upper half-plane, we see that any single-valued branch of $f(z)$ in the lower half-plane must have a simple pole at $\bar{\beta}$. Therefore at $\bar{\beta}$, $f(z)$ has an expansion of the form

$$\frac{f''(z)}{f'(z)} = -\frac{2}{z - \bar{\beta}} + \cdots.$$

It follows by the same argument as in the proof of Theorem 9.6 that

$$\frac{f''(z)}{f'(z)} - \sum_{j=1}^{n} \frac{\alpha_j - 1}{z - a_j} - \frac{2}{z - \beta} - \frac{2}{z - \bar{\beta}} = 0. \tag{9.5}$$

Integrating (9.5), we obtain

$$f'(z) = C\prod_{j=1}^{n}(z - a_j)^{\alpha_j - 1}(z - \beta)^{-2}(z - \bar{\beta})^{-2},$$

and another integration gives the required representation (9.4).

THEOREM 9.10. *Let G be a polygon in the extended plane which contains ∞ as an interior point. Then the sum of the vertex angles of G equals $(n + 2)\pi$.*

Proof. This time we have

$$\sum_{j=1}^{n}(\alpha_j - 1) - 4 + 2 = 0,$$

and hence

$$\sum_{j=1}^{n}\alpha_j = n + 2.$$

Formulas (9.1) and (9.4) were derived under the assumption that the inverse images $a_1, \ldots, a_n$ of the vertices $w_1, \ldots, w_n$ are all finite. We now relax this condition:

THEOREM 9.11. *If $a_n = \infty$, then formulas* (9.1) *and* (9.4) *are replaced by*

$$w = f(z) = C\int_{z_0}^{z}\prod_{j=1}^{n-1}(t - a_j)^{\alpha_j - 1}\,dt + C_1 \tag{9.1'}$$

and

$$w = f(z) = C\int_{z_0}^{z}\frac{\prod_{j=1}^{n-1}(t - a_j)^{\alpha_j - 1}}{(t - \beta)^2(t - \bar{\beta})^2}\,dt + C_1, \tag{9.4'}$$

i.e., the factor $(z - a_n)^{\alpha_n - 1}$ disappears from the product

$$\prod_{j=1}^{n}(z - a_j)^{\alpha_j - 1}.$$

Proof. The transformation

$$z' = \frac{1}{a_n - z}, \tag{9.6}$$

or equivalently

$$z = a_n - \frac{1}{z'},$$

maps the upper half-plane onto itself and transforms (9.1) into

$$w = f(z) = C\int_{z_0}^{a_n - \frac{1}{z'}}\prod_{j=1}^{n}(t - a_j)^{\alpha_j - 1}\,dt + C'. \tag{9.7}$$

After the change of variable

$$t = a_n - \frac{1}{t'},$$

(9.7) becomes

$$w = f(z) = C \int_{\frac{1}{a_n - z_0}}^{z'} \prod_{j=1}^{n} [(a_n - a_j)t' - 1]^{\alpha_j - 1} t'^{-\sum_{j=1}^{n} (\alpha_j - 1)} \frac{dt'}{t'^2} + C_1$$

or

$$w = f(z) = C_1 \int_{z_0'}^{z'} \prod_{j=1}^{n-1} (t' - a_j')^{\alpha_j - 1}\, dt' + C_1, \tag{9.8}$$

where

$$z_0' = \frac{1}{a_n - z_0}, \qquad a_j' = \frac{1}{a_n - a_j}$$

as in (9.6) and we use the fact that

$$\sum_{j=1}^{n} (\alpha_j - 1) = -2. \tag{9.9}$$

Dropping the primes, we find that (9.8) coincides with (9.1′). Formula (9.4′) is proved in the same way, except that now the right-hand side of (9.9) equals $+2$.

Finally we map the unit disk onto a polygon:

THEOREM 9.12. *If* Π_U: Im $z > 0$ *is replaced by* K: $|z| < 1$, *then* (9.1) *remains the same, except that* $z_0 \in \bar{K}$, $z \in K$ *and the inverse images* $a_1, \ldots, a_n$ *of the vertices of the polygon are now of the form* $a_1 = e^{i\theta_1}, \ldots, a_n = e^{i\theta_n}$ *where* $0 \leqslant \theta_j < 2\pi$ $(j = 1, \ldots, n)$ *and* $\theta_1 < \cdots < \theta_n$. *Formula* (9.4) *becomes*

$$w = f(z) = C \int_{z_0}^{z} \prod_{j=1}^{n} (t - a_j)^{\alpha_j - 1} \frac{dt}{t^2} + C_1, \tag{9.10}$$

where $z = 0$ *is the inverse image of the point at infinity.*

Proof. The transformation

$$z' = \frac{z - \zeta}{z - \bar{\zeta}} \qquad (\text{Im } \zeta > 0) \tag{9.11}$$

maps Π_U onto K and transforms (9.11) into

$$w = f(z) = C \int_{z_0}^{(\zeta - \bar{\zeta}z')/(1 - z')} \prod_{j=1}^{n} (t - a_j)^{\alpha_j - 1}\, dt + C_1. \tag{9.12}$$

After the change of variable

$$t = \frac{\zeta - \bar{\zeta}t'}{1 - t'},$$

(9.12) becomes

$$w = f(z) = C\int_{z_0'}^{z'} \prod_{j=1}^{n} [\zeta - \bar{\zeta}t' - a_j(1 - t')]^{\alpha_j - 1}\left(\frac{1}{1 - t'}\right)^{\sum_{j=1}^{n}(\alpha_j - 1)} \times \frac{\zeta - \bar{\zeta}}{(1 - t')^2}\, dt' + C_1, \qquad (9.13)$$

where $z_0' \in \bar{K}$, $z' \in K$ are the images of $z_0 \in \overline{\Pi}_U$, $z \in \Pi_U$ under the transformation (9.11). In terms of the points

$$a_j' = \frac{a_j - \zeta}{a_j - \bar{\zeta}} \qquad (j = 1, \ldots, n),$$

which are clearly of the required form, (9.13) can be written as

$$w = f(z) = C'\int_{z_0'}^{z'} \prod_{j=1}^{n} (t' - a_j')^{\alpha_j - 1}\, dt' + C_1. \qquad (9.14)$$

Dropping the primes, we find that (9.14) coincides with (9.1).

In the case of a polygon with $w = \infty$ as an interior point, we map $K: |z| < 1$ onto the polygon in such a way that $z = 0$ goes into the point at infinity. Since (9.4) carries $z = \beta$ into $w = \infty$, the appropriate form of (9.11) is now

$$z' = \frac{z - \beta}{z - \bar{\beta}}.$$

Applying this transformation to (9.4), we obtain (9.10) after making some elementary calculations and using the formula

$$\sum_{j=1}^{n} (\alpha_j - 1) = 2.$$

48. Examples

We now give some explicit calculations illustrating the theory of the preceding section.

Example 1. *Map the upper half-plane onto the domain shown in Figure* 9.2(e).

Let the two rays forming the boundary of the domain have equations $w = u_1 + t$ $(t \geqslant 0)$ and $w = u_2 + ih - t$ $(t \geqslant 0)$. Then the vertices of the polygon are ∞, u_1, ∞, $u_2 + ih$ (in order) and the vertex angles are $-\pi$, 2π,

$-\pi$, 2π. Suppose we require the inverse images of the vertices to be the points 0, 1, ∞, a of the real axis, where $a < 0$. Then, according to (9.1′), the required mapping is given by

$$w = C\int_{z_0}^{z} t^{-2}(t-1)(t-a)\,dt + C_1 = Cz - (a+1)C\,\mathrm{Ln}\,z - aCz^{-1} + C_1',$$

where we need only choose one of the branches of $\mathrm{Ln}\,z$ in the upper half-plane, say $\ln z$ (choosing another branch merely changes the value of C_1'). Therefore

$$w = Cz - (a+1)C\ln z - aCz^{-1} + C',$$

where we drop the subscript on C_1'. It is easy to see that C and C' are real. In fact, as z traverses the interval $(1, \infty)$, w describes side II of the polygon, which lies on the real axis [see Figure 9.2(e)]. Choosing two points z', $z'' \in (1, \infty)$ with images $w' = f(z')$, $w'' = f(z'')$, we obtain the following system of equations determining C and C', whose solution is obviously real:

$$\begin{aligned} w' &= [z' - (a+1)\ln z' - az'^{-1}]C + C', \\ w'' &= [z'' - (a+1)\ln z'' - az''^{-1}]C + C'. \end{aligned}$$

To actually calculate C, C' and a, we use the equations

$$\begin{aligned} u_1 &= C - aC + C', \\ u_2 + ih &= Ca - (a+1)C\ln a - C + C', \end{aligned}$$

which imply

$$\begin{aligned} u_1 &= C - aC + C', \\ u_2 &= Ca - (a+1)C\ln|a| - C + C', \\ h &= -(a+1)\pi, \end{aligned}$$

since $\ln a = \ln|a| + i\pi$. It follows that

$$\begin{aligned} C &= -\frac{h}{\pi(1+a)}, \\ C' &= u_1 + \frac{h}{\pi}\frac{1-a}{1+a}, \\ 2\frac{h}{\pi}\frac{1-a}{1+a} &+ \frac{h}{\pi}\ln|a| = u_2 - u_1. \end{aligned}$$

Solving the last equation for a, we can then determine C and C', thereby completing the calculation.

Example 2. *Map the upper half-plane onto the domain shown in Figure* 9.2(f).

The two rays forming the boundary of the domain now have equations $w = u_1 - t\,(t \geqslant 0)$ and $w = u_2 + ih - t\,(t \geqslant 0)$. Moreover, the vertices and vertex angles are now u_1, ∞, $u_2 + ih$, ∞ and 2π, -2π, 2π, 0. Suppose we require the inverse images of the vertices to be the points b, ∞, -1 and 0 of

the real axis, where $b > 0$. Then, according to (9.1′), the required mapping is given by

$$w = C\int_{z_0}^{z} (t-b)(t+1)^{-1}t^{-1}\,dt + C_1 = C[\tfrac{1}{2}z^2 - (b-1)z - b\ln z] + C',$$

where the constants C and C' are again real. To determine C, C' and b, we use the equations

$$\begin{aligned} u_1 &= C[\tfrac{1}{2}b^2 - (b-1)b - b\ln b] + C', \\ u_2 + ih &= C[1 + (b-1) - b\ln(-1)] + C', \end{aligned}$$

which imply

$$\begin{aligned} C &= -\frac{h}{b\pi}, \\ C' &= u_2 + \frac{h(2b-1)}{2b\pi}, \\ \frac{bh}{2\pi} - \frac{h}{2b\pi} + \frac{h}{\pi}\ln b &= u_1 - u_2. \end{aligned}$$

Solving the last equation for b, we can then determine C and C', thereby completing the calculation. It should be noted that there is no need to prove the existence of a root of the transcendental equation for $b > 0$ (or of the corresponding equation for $a < 0$ in Example 1), since the existence of the root is implied by the existence of the mapping $w = f(z)$ itself.

Example 3. *Map the unit disk* $|z| < 1$ *onto the domain shown in Figure* 9.5(a), *whose boundary consists of* $2n$ *line segments joining the origin to the points*

$$1, \quad e^{\pi i/n}, \ldots, \quad e^{(2n-1)\pi i/2n}.$$

Since each sequence is traversed twice, the polygon in the w-plane has $4n$ sides, with vertices

$$1, \quad 0, \quad e^{\pi i/n}, \quad 0, \ldots, \quad e^{(2n-1)\pi i/2n}, \quad 0$$

and vertex angles

$$2\pi, \quad \frac{\pi}{n}, \quad 2\pi, \quad \frac{\pi}{n}, \ldots, \quad 2\pi, \quad \frac{\pi}{n}.$$

The corresponding inverse images on the unit circle $|z| = 1$ must be the vertices of a regular $4n$-gon. In fact, suppose $w = f(z)$ maps the circular sector $A'B'O_1'$ of vertex angle $\pi/2n$ in the z-plane [see Figure 9.5(b)] conformally onto the angle AOB in the w-plane [see Figure 9.5(a)], in such a way that the points 1, 0, $e^{-\pi i/2n}$ go into the points 1, ∞, 0. Then, applying the symmetry principle to the function $f(z)$, the domain $A'B'O_1'$ and the segment $B'O_1'$, we find that the same function (after analytic continuation) maps the sector $A'B'C'$ (whose vertex angle is twice as large as that of $A'B'O_1'$) conformally onto the angle AOC in such a way that the radii $B'A'$

and $B'C'$ go into the rays AB and CB, while the arc $A'O_1'C'$ goes into the polygonal curve AOC with two segments. Applying the symmetry principle to the function $f(z)$, the domain $A'B'C'$ and the segment $B'C'$, and then once again to the same function, the domain $C'B'D'$ and the segment $B'D'$, and so on, we find that $f(z)$ can be continued analytically into the whole unit disk (there is a pole at the origin), which it maps conformally onto the given $4n$-gon. Under this mapping, the inverse images of the $4n$-gon in the w-plane are the vertices of a regular $4n$-gon inscribed in the unit circle $|z| = 1$.

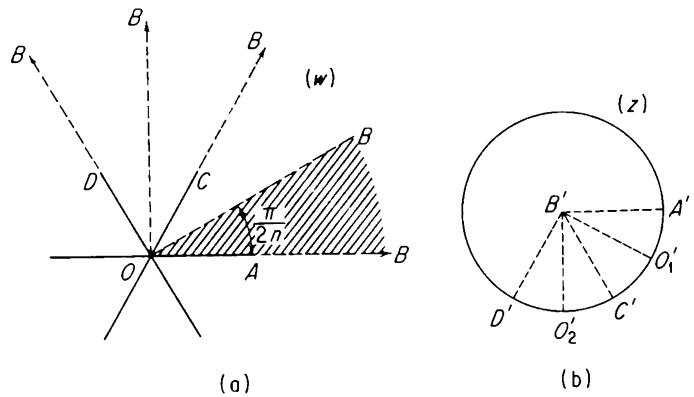


FIGURE 9.5

It follows from the above considerations that the required mapping $w = f(z)$ can be found directly, and this will be done below. However, to illustrate the technique of Sec. 47, we now use Theorem 9.12 instead. Since the inverse images of the vertices of the $4n$-gon are

$$1, \quad e^{-\pi i/2n}, \ldots, \quad e^{-(2n-2)\pi i/2n}, \quad e^{-(2n-1)\pi i/2n}$$

in order (starting from A), formula (9.10) becomes

$$\begin{aligned} w = f(z) &= C\int_1^z (t-1)(t-e^{-\pi i/2n})^{(1/n)-1}\cdots(t-e^{-(2n-2)\pi i/2n}) \\ &\qquad \times (t-e^{-(2n-1)\pi i/2n})^{(1/n)-1}\frac{dt}{t^2} + C_1 \\ &= C\int_1^z (t^{2n}-1)(t^{2n}+1)^{(1/n)-1}\frac{dt}{t^2} + C_1 \\ &= C\int_1^z (t^n + t^{-n})^{(1/n)-1}(t^{n-1} - t^{-n-1})\,dt + C_1 \\ &= C(z^n + z^{-n})^{1/n} + C_1. \end{aligned}$$

To determine the constants C and C_1, we set $z = 1$ and $z = e^{-\pi i/2n}$, obtaining

$$1 = 2^{1/2}C + C_1, \qquad 0 = C_1,$$

so that finally

$$w = f(z) = [\tfrac{1}{2}(z^n + z^{-n})]^{1/n}. \tag{9.15}$$

The same problem can also be solved directly by noting that the effect of consecutively applying the following three transformations is equivalent to formula (9.15):

1. The mapping $\eta = w^n$ which transforms the w-plane into an n-sheeted Riemann surface and the $4n$-gon into a domain D with boundary consisting of $2n$ line segments of unit length whose projections onto the real axis are alternately the segments $[0, 1]$ and $[-1, 0]$;
2. The mapping $\eta = \frac{1}{2}(\eta_1 + \eta_1^{-1})$ which transforms D into an "n-sheeted unit disk," provided we choose the branch of $\eta_1 = \eta + \sqrt{\eta^2 - 1}$ which vanishes at infinity;
3. The mapping $z = \eta_1^{1/n}$ which transforms the n-sheeted disk back into a single-sheeted disk.

49. The Modular Function. Picard's First Theorem

We now use the symmetry principle to construct an important function with many theoretical implications. Let A, B and C be three "equally spaced" points of the unit circle $\gamma: |z| = 1$, and draw the circular arcs AB, BC and CA, each orthogonal to γ, as shown in Figure 9.6(a). These arcs

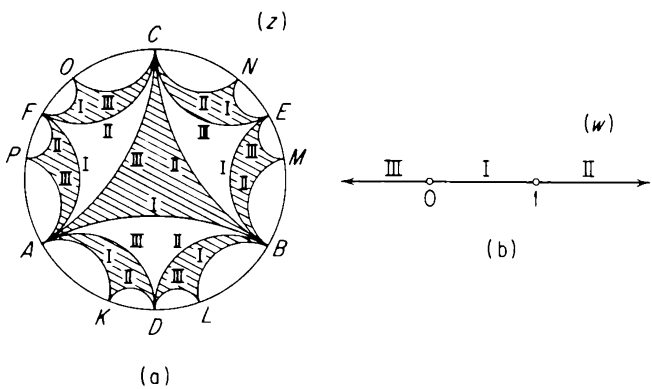


FIGURE 9.6

form the boundary of a triangular domain G, whose vertex angles are all zero. Let $w = \chi(z)$ be the function mapping G conformally onto the upper half-plane $\operatorname{Im} w > 0$ in such a way that the points A, B and C go into the points 0, 1 and ∞. Then the arcs AB, BC and CA go into the intervals $[0, 1]$, $[1, \infty]$ and $[-\infty, 0]$ of the real axis, which we denote by I, II and III, as indicated in Figure 9.6(b).

It follows from the symmetry principle, applied to the domain G and the function $\chi(z)$, that $\chi(z)$ can be continued analytically across each of the arcs AB, BC and CA into the triangles ABD, BCE and ACF (whose vertex angles

are all zero). Moreover, after the continuation, $\chi(z)$ maps each of these triangles onto the lower half-plane. Applying the symmetry principle again to $\chi(z)$ and each of the triangles ABD, BCE and ACF, we find that $\chi(z)$ can be continued analytically across the arcs AD, DB, BE, EC, CF and AC into the triangles ADK, DBL, BEM, ECN, CFO and FAP (with zero vertex angles), each of which is mapped onto the upper half-plane. Repeating this construction indefinitely, we find that at each stage $\chi(z)$ is continued into a polygonal domain whose sides are circular arcs orthogonal to γ. In fact, starting from the triangle ABC, we continue $\chi(z)$ into the hexagon $ADBECF$, then into the dodecagon $AKDLMBENCOFP$, and so on.

THEOREM 9.13. *The construction just described continues $\chi(z)$ into the whole unit disk $K: |z| < 1$. Moreover, $\chi(z)$ is single-valued on K.*

Proof. Given any point $z_0 \in K$, we must show that z_0 belongs to the polygon obtained at some stage of the construction. Clearly, it is sufficient to prove that every point of γ is a limit point of vertices of the polygons, since then the lengths of the arcs of γ "supporting" the sides of the polygons go to zero as the number of sides becomes infinite,[14] which in turn implies that the polygon contains z_0 if it has sufficiently many sides (why?).

Thus, suppose there is a point of γ which is not a limit point of vertices of the polygons. Then there exists a nested sequence $\{\sigma_n\}$ of arcs of γ, joining adjacent vertices of the polygons, which approach an arc $\sigma \subset \gamma$ with no vertices of the polygons as interior points. Let τ be the circular arc which has the same end points as σ and is orthogonal to γ, and let A', B', C' be the points symmetric to A, B, C with respect to τ (see Figure 9.7). Reflecting the polygons in the sides supported by σ_n, we find that the images of A, B and C are arbitrarily close to A', B' and C' for sufficiently large n. Therefore, starting from some value of n, at least one vertex of every polygon is an interior point of σ (being distinct from its end points). This contradiction proves that $\chi(z)$ can be continued into the whole disk K. The fact that $\chi(z)$ is single-valued on K follows from the monodromy theorem.[15]

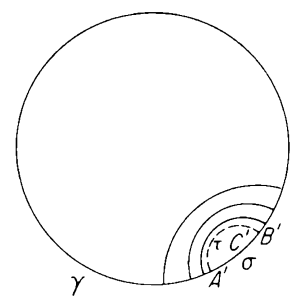


FIGURE 9.7

Because of the way it is constructed, the function $\chi(z)$ takes every complex value except the three values 0, 1 and ∞ at infinitely many points of the unit

[14] Of course, we always have in mind the smaller of the two arcs with given end points.

[15] The single-valued character of $\chi(z)$ is also an immediate consequence of the fact that at each stage of the construction, the polygon G being reflected lies outside the circle containing its side. Therefore G and its image G^* under reflection are disjoint.

disk. In fact, every shaded triangle is mapped onto the upper half-plane and every unshaded triangle onto the lower half-plane, while the sides labelled I, II and III are mapped onto the intervals $[0, 1]$, $[1, \infty]$ and $[-\infty, 0]$ of the real axis. The function $w = \chi(z)$ determines a Riemann surface which can be regarded as made up of infinitely many replicas of the upper and lower half-planes, where every upper (or lower) half-plane is joined to three lower (or upper) half-planes along the segments I, II and III of the real axis. On this surface, the inverse of the function $w = \chi(z)$ is single-valued, analytic and bounded (its values belong to the unit disk). Correspondingly, the function

$$\omega(w) = \chi^{-1}(w) \tag{9.16}$$

is multiple-valued and can be continued analytically along any continuous curve which does not pass through the points 0, 1 and ∞. Among the values taken by $\omega(w)$, there is one and only one, called the *principal value* of $\omega(w)$, which is taken in the closed triangle ABC or the open triangle ADB.

If we map the unit disk $|z| < 1$ onto the upper half-plane $\operatorname{Im} t > 0$ in such a way that the points A, B and C go into the points 0, 1 and ∞, then the

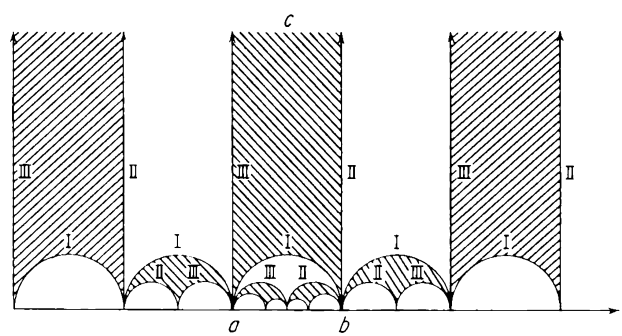


FIGURE 9.8

triangle ABC with zero angles goes into the triangle abc with zeros angles shown in Figure 9.8. At the same time, the function $w = \chi(z)$ is transformed into a function $w = \lambda(t)$ called the *modular function*. The modular function is single-valued and analytic in the upper half-plane, and is completely characterized by the fact that it maps the triangle abc with zero angles and vertices 0, 1, ∞ onto the upper half-plane in such a way that the vertices 0, 1, ∞ go into the points 0, 1, ∞ of the real axis. All the properties of the function $w = \lambda(t)$ are immediate consequences of those of $w = \chi(z)$. For example, the modular function takes every complex value except the three values 0, 1 and ∞ at infinitely many points of the upper half-plane. Moreover, every shaded triangle is mapped onto the upper half-plane and every unshaded triangle onto the lower half-plane, while the sides labelled I, II and III are mapped onto the intervals $[0, 1]$, $[1, \infty]$ and $[-\infty, 0]$ of the real axis. The Riemann (covering) surface of the function $w = \lambda(t)$ is equivalent

to that of the function $w = \chi(z)$,[16] and is called the *Riemann surface of the modular function.*

Turning to the implications of the above theory, we now prove a result which is already familiar for the special case of entire functions of finite order (see Theorem II.9.11):

THEOREM 9.14 (*Picard's first theorem*).[17] *An entire function* $f(z) \not\equiv$ const *takes every finite complex value with one possible exception,*

Proof. Suppose $f(z)$ fails to take two distinct finite values a and b. Then

$$\varphi(z) = \frac{f(z) - a}{b - a}$$

is an entire function which takes neither of the values 0 and 1. Therefore the range of $\varphi(z)$ is contained in the domain of the multiple-valued function (9.16). Choosing any single-valued branch of $\omega(w)$ in a neighborhood of the point $w_0 = \varphi(0)$, we obtain an element of the analytic function $\omega[\varphi(z)]$. This element can be continued analytically along every curve in the finite plane, and hence, by the monodromy theorem and its corollary, defines a function which is single-valued and analytic in the finite plane, i.e., an entire function. Moreover $\omega[\varphi(z)]$ is bounded, since its values all lie in the unit disk. Therefore, by Liouville's theorem, $\omega[\varphi(z)] \equiv$ const, which is possible only if $\varphi(z) \equiv$ const, or equivalently $f(z) \equiv$ const, contrary to hypothesis. This contradiction proves the theorem.

COROLLARY. *A meromorphic function* $f(z) \not\equiv$ const *takes every complex value with two possible exceptions.*

Proof. Suppose $f(z)$ fails to take three distinct values a, b and c. If one of these values, say c, is infinite, then $f(z) \not\equiv$ const is an entire function which fails to take two finite values, and this is impossible, as just shown. Therefore a, b and c are finite. Consider the function

$$\varphi(z) = \frac{1}{f(z) - c} \not\equiv \text{const.}$$

Since $\varphi(z)$ is meromorphic and does not become infinite in the finite plane, $\varphi(z)$ must be entire. But $\varphi(z)$ cannot take the two finite values

$$\frac{1}{a - c}, \qquad \frac{1}{b - c}.$$

This again contradicts Theorem 9.14, thereby proving the corollary.

---

[16] Since one is carried into the other by the Möbius transformation mapping the unit disk onto the upper half-plane (recall Definition 3, p. 238).

[17] Also known as *Picard's little theorem.* A stronger version of Theorem 9.14 will be proved later (see Theorem 9.16, Corollary 1), allowing us to add the phrase "at infinitely many points" to the statement of Theorem 9.14, provided $f(z)$ is transcendental.

Example. The meromorphic function e^z has two exceptional values 0 and ∞, while the meromorphic function $\tan z$ has two exceptional values $\pm i$ (recall p. I.205).

50. Normal Families of Analytic Functions

It will be recalled from Sec. I.86 that a family $\mathscr{F}$ of (single-valued) functions, all analytic on the same domain G, is said to be *compact* (on G) if every sequence $\{f_n(z)\}$ of functions $f_n(z) \in \mathscr{F}$ contains a subsequence $\{f_{n_k}(z)\}$ which converges uniformly inside G.[18] We now generalize this concept somewhat, by allowing the limit function to be identically infinite:

DEFINITION. *A sequence of functions $\{f_n(z)\}$, all analytic on a domain G, is said to converge uniformly to infinity [or to the limit function $f(z) \equiv \infty$] inside G if, given any closed bounded set $F \subset G$ and any number $M > 0$, there exists an integer $N = N(F, M) > 0$ such that $|f_n(z)| > M$ for all $n > N$ and $z \in F$. A family $\mathscr{F}$ of functions, all analytic on a domain G, is said to be normal (on G) if every sequence $\{f_n(z)\}$ of functions $f_n(z) \in \mathscr{F}$ contains a subsequence $\{f_{n_k}(z)\}$ which converges uniformly inside G either to a finite function (which is necessarily analytic on G) or to the function $f(z) \equiv \infty$.*

Remark. Thus $\mathscr{F}$ is normal if $\mathscr{F}$ is compact, and in particular, according to Theorem I.17.17, $\mathscr{F}$ is normal if $\mathscr{F}$ is uniformly bounded inside G. On the other hand, the sequence $\{z + n\}$ converges uniformly to infinity on the unit disk K and hence is normal, but $\{z + n\}$ cannot be compact, since it is not uniformly bounded inside K (invoke Theorem I.17.17 again).

Using the properties of the function $\omega(w)$, we can now prove

THEOREM 9.15 (*Montel's normality criterion*). *Given a family $\mathscr{F}$ of functions analytic on a domain G, suppose no function in $\mathscr{F}$ takes either of two distinct finite values a and b at a point of G. Then $\mathscr{F}$ is normal on G.*

Proof. Recalling the considerations on p. I.416, we see that it is sufficient to prove that $\mathscr{F}$ is normal on every disk contained in G, i.e., there is no loss of generality in assuming from the outset that G is a disk. Moreover, replacing every function $f(z) \in \mathscr{F}$ by

$$\varphi(z) = \frac{f(z) - a}{b - a},$$

we obtain a family of functions $\mathscr{F}'$ none of which takes the values 0 and 1 in the disk G. Clearly, it is enough to prove that $\mathscr{F}'$ is normal (since normality of $\mathscr{F}'$ implies that of $\mathscr{F}$).

---

[18] I.e., on every closed bounded subset of G (cf. footnote 21, p. I.415).

Thus let $\{\varphi_n(z)\}$ be any sequence of functions in $\mathscr{E}'$, and consider the numerical sequence $\{\varphi_n(z_0)\}$, where z_0 is any point of G, Suppose first that $\{\varphi_n(z_0)\}$ has a limit point different from 0, 1 and ∞ (the other alternative will be considered presently). Then, going over to a subsequence if necessary, we have

$$\lim_{n\to\infty} \varphi_n(z_0) = \alpha \neq 0, 1, \infty. \tag{9.17}$$

Fixing a neighborhood $\mathscr{N}(\alpha)$ so small that it does not contain the points 0, 1 and ∞, we choose a single-valued branch of $\omega(w)$ on $\mathscr{N}(\alpha)$, e.g., by requiring that $\beta = \omega(\alpha)$ be the principal value of $\omega(\alpha)$.[19] Since $\varphi_n(z_0) \in \mathscr{N}(\alpha)$ for sufficiently large n, this allows us to define elements of the form $\{\mathscr{N}(\alpha), \omega\varphi_n(z)\}$,[20] where the values of ω belong to the branch just selected. Each of these elements can be continued analytically into the whole disk G, since the $\varphi_n(z)$ are analytic on G and their values all belong to the domain of $\omega(w)$. As a result of the continuation, we obtain a sequence of single-valued functions $\omega\varphi_n(z)$, which are analytic and satisfy the condition

$$\lim_{n\to\infty} \omega\varphi_n(z_0) = \beta.$$

Moreover, the values of every function $\omega\varphi_n(z)$ belong to the unit disk. Therefore, by Theorem I.17.17, the sequence $\{\omega\varphi_n(z)\}$ is compact, and hence contains a subsequence $\{\omega\varphi_{n_k}(z)\}$ which converges uniformly inside G. Setting

$$\Omega(z) = \lim_{k\to\infty} \omega\varphi_{n_k}(z),$$

we find that

$$\Omega(z_0) = \beta,$$

In particular, since $|\beta| < 1$, $\Omega(z)$ cannot be a constant of absolute value unity. Moreover

$$|\omega\varphi_{n_k}(z)| < 1 \qquad (z \in G),$$

and hence

$$|\Omega(z)| \leqslant 1 \qquad (z \in G).$$

Therefore

$$|\Omega(z)| < 1 \qquad (z \in G)$$

by the maximum modulus principle, since if $\Omega(z) = 1$ at any point of G, then $|\Omega(z)| \equiv 1$, which, as just noted, is impossible. Let $\overline{K}$ be a closed disk contained in G. Then the function $t = \chi(z)$ maps $\overline{K}$ onto a closed domain $\overline{D}$ contained in the unit disk $|z| < 1$. Moreover, the function

---

[19] I.e., the value belonging to the quadrilateral $ADBC$ shown in Figure 9.6, not including the sides AD and DB (recall p. 338).

[20] For simplicity, we write $\omega\varphi_n(z)$ instead of $\omega[\varphi_n(z)]$, $\chi\omega\varphi_{n_k}(z)$ instead of $\chi\{\omega[\varphi_{n_k}(z)]\}$, etc.

$w = \chi(t)$ is bounded on $\bar{D}$, and hence, by Theorem I.15.9, the sequence

$$\{\chi\omega\varphi_{n_k}(z)\} = \{\varphi_{n_k}(z)\}$$

converges uniformly on $\bar{K}$ to the function $\chi\Omega(z)$. Therefore, in the case where $\{\varphi_n(z_0)\}$ has a limit point different from 0, 1 and ∞, the sequence $\{\varphi_n(z)\}$ contains a subsequence $\{\varphi_{n_k}(z)\}$ which converges uniformly inside G.

Next we turn to the case where $\{\varphi_n(z_0)\}$ has no limit points other than 0, 1 and ∞. If 1 is a limit point of $\{\varphi_n(z_0)\}$, there is no loss of generality in assuming that

$$\lim_{n\to\infty} \varphi_n(z_0) = 1. \tag{9.18}$$

Consider the function

$$\psi_n(z) = \frac{\operatorname{Ln} \varphi_n(z) + 2\pi i}{4\pi i},$$

where $\operatorname{Ln} \varphi_n(z)$ is assigned its principal value at z_0, so that

$$\lim_{n\to\infty} \operatorname{Ln} \varphi_n(z_0) = 0.$$

Continuing the corresponding elements of $\operatorname{Ln} \varphi_n(z)$ into the disk G,[21] we obtain functions $\psi_n(z)$ which are single-valued and analytic on G. Since $\operatorname{Ln} \varphi_n(z) \neq \pm 2\pi i, \infty$ in G, the functions $\psi_n(z)$ cannot take the values 0, 1 and ∞, and moreover

$$\lim_{n\to\infty} \psi_n(z_0) = \tfrac{1}{2} \neq 0, 1, \infty.$$

This allows us to apply the same argument to the sequence $\{\psi_n(z)\}$ as applied to $\{\varphi_n(z)\}$ in the case where $\varphi_n(z_0) \to \alpha \neq 0, 1, \infty$. Hence there exists a subsequence $\psi_{n_k}(z)$ which converges uniformly inside G to a function $\Psi(z)$. Since none of the functions $\psi_{n_k}(z)$ can equal $\frac{1}{2}$ in G, it follows from Hurwitz's theorem (Theorem II.2.5) that the same is true of the limit function $\Psi(z)$, unless $\Psi(z)$ reduces to a constant. But

$$\Psi(z_0) = \lim_{k\to\infty} \psi_{n_k}(z_0) = \tfrac{1}{2},$$

and hence

$$\Psi(z) \equiv \tfrac{1}{2}.$$

Therefore the branches of $\operatorname{Ln} \varphi_{n_k}(z)$ selected above converge uniformly to 0 inside G, i.e., the sequence $\{\varphi_{n_k}(z)\}$ converges uniformly to 1 inside G.

Finally, suppose that instead of (9.18), we have

$$\lim_{n\to\infty} \varphi_n(z_0) = 0 \tag{9.19}$$

or

$$\lim_{n\to\infty} \varphi_n(z_0) = \infty. \tag{9.20}$$

---

[21] The continuation is possible since $\varphi_n(z) \neq 0, \infty$ in G.

Then we consider the sequences

$$\{1 - \varphi_n(z)\} \tag{9.21}$$

and

$$\left\{1 - \frac{1}{\varphi_n(z)}\right\}, \tag{9.22}$$

each consisting of functions which are analytic on G and do not take the values 0, 1 and ∞. The sequences (9.21) and (9.22) both converge to 1 at z_0, and hence, by the argument just given, $\{\varphi_n(z)\}$ contains a subsequence converging uniformly inside G to 0 if (9.21) holds or to ∞ if (9.22) holds. The proof of the fact that $\mathscr{F}'$ (and hence $\mathscr{F}$) is normal on G is now complete, since together (9.17), (9.18), (9.19) and (9.20) account for all the possibilities.

51. Picard's Second Theorem. Julia Directions

Using the normality criterion, we can now prove one of the key theorems of analytic function theory:

THEOREM 9.16 (*Picard's second theorem*).[22] *In any deleted neighborhood of an essential singular point z_0, an analytic function $f(z)$ takes every finite value with one possible exception.*

Proof. There is no loss of generality in assuming that $z_0 = 0$, since we can always arrive at this case by making a preliminary transformation of the form $Z = z - z_0$ or $Z = 1/z$. Suppose $f(z)$ has two exceptional values a and b in some neighborhood $|z| < R$ of the point $z = 0$. Introducing the annular domains

$$D_0: \frac{R}{2} < |z| < R, \quad D_1: \frac{R}{2^2} < |z| < \frac{R}{2}, \ldots, D_n: \frac{R}{2^{n+1}} < |z| < \frac{R}{2^n}, \ldots,$$

we construct the sequence of functions

$$f_1(z) = f\left(\frac{z}{2}\right), \ldots, f_n(z) = f\left(\frac{z}{2^n}\right), \ldots$$

The function $f_n(z)$ takes the same values in the annulus D_0 as the function $f(z)$ takes in D_n, and hence $\{f_n(z)\}$ is a sequence of functions analytic on D_0 none of which takes the values a and b. Therefore, by Theorem 9.15, $\{f_n(z)\}$ is a normal family, i.e., $\{f_n(z)\}$ contains a subsequence $\{f_{n_k}(z)\}$ converging uniformly inside G to a limit function $F(z)$, which may be

[22] Also known as *Picard's great theorem.* The implications of this theorem have already been discussed on p. II.21.

identically infinite. First suppose $F(z) \not\equiv \infty$. Then $F(z)$ is bounded on the circle

$$\gamma: |z| = \rho \qquad \left(\frac{R}{2} < \rho < R\right).$$

It follows that the functions $f_{n_k}(z)$ are uniformly bounded on γ, which in turn implies

$$|f(z)| \leqslant M \qquad \text{for} \quad |z| = \frac{\rho}{2^{n_k}} \qquad (k = 1, 2, \ldots).$$

Therefore, by the maximum modulus principle,

$$|f(z)| \leqslant M \qquad (0 < |z| \leqslant \rho).$$

But this is impossible, since $z = 0$ is an essential singular point of $f(z)$.

If $F(z) \equiv \infty$, the sequence of analytic functions

$$\varphi_{n_k}(z) = \frac{1}{f_{n_k}(z) - a}$$

converges uniformly to zero inside D_0, and hence by the argument just given, the function

$$\varphi(z) = \frac{1}{f(z) - a}$$

is bounded on a set of the form $0 < |z| \leqslant \rho$. But then $z = 0$ is a regular point of $\varphi(z)$, and hence $z = 0$ cannot be an essential singular point of $f(z)$, contrary to hypothesis. This contradiction completes the proof.

COROLLARY 1. *An entire transcendental function* $f(z) \not\equiv$ const *takes every finite value with one possible exception at infinitely many points.*

Proof. If $f(z)$ takes two distinct finite values a and b at only finitely many points, then $f(z)$ fails to take the values a and b in some deleted neighborhood of the point at infinity. But this contradicts Theorem 9.16, since $z = \infty$ is an essential singular point of $f(z)$.

COROLLARY 2. *A meromorphic function* $f(z) \not\equiv$ const *with an essential singular point at* $z = \infty$ *takes every finite value with two possible exceptions at infinitely many points.*

Proof. Recall the proof of the corollary to Theorem 9.14.

Next we prove a closely related result which will be used below to deduce information on the distribution of A-points of $f(z)$ near an essential singular point:

LEMMA. *Let* $f(z)$ *be analytic on* $0 < |z| < R$, *with an essential singular point at* $z = 0$. *Then the sequence*

$$f_1(z) = f\left(\frac{z}{2}\right), \ldots, f_n(z) = f\left(\frac{z}{2^n}\right), \ldots \tag{9.23}$$

cannot be normal on the annulus

$$D_0' : \frac{R}{2^4} < |z| < R.$$

Proof. Suppose $\{f_n(z)\}$ is normal on D_0'. Then $\{f_n(z)\}$ contains a subsequence $\{f_{n_k}(z)\}$ converging uniformly inside D_0', in particular converging uniformly on the closed annulus

$$\bar{\Delta} : \frac{R}{2^3} \leqslant |z| \leqslant \frac{R}{2}.$$

The limit function $F(z)$ of $\{f_{n_k}(z)\}$ must be identically infinite, since otherwise, just as in the proof of Theorem 9.16, $f(z)$ would be bounded on a set of the form $0 < |z| \leqslant \rho$, which is impossible. But if $\{f_{n_k}(z)\}$ converges uniformly to infinity on $\bar{\Delta}$, then so does the original sequence $\{f_n(z)\}$, since otherwise there would exist a number $M > 0$, a sequence of points $z_k \in \bar{\Delta}$ and a subsequence $\{f_{m_k}(z)\}$ such that

$$|f_{m_k}(z_k)| \leqslant M \qquad (k = 1, 2, \ldots),$$

contrary to the fact that $\{f_{m_k}(z)\}$ must contain another subsequence $\{f_{m_k'}(z)\}$ converging uniformly to infinity on $\bar{\Delta}$. Therefore, given any $M > 0$, there is an integer $\nu = \nu(M) > 0$ such that

$$|f_n(z)| > M$$

for all $n > \nu$ and $z \in \bar{\Delta}$. As z varies over the domain $\bar{\Delta}$, the point $z/2^n$ varies over the closed annulus

$$\bar{\Delta}_n : \frac{R}{2^{n+3}} \leqslant |z| \leqslant \frac{R}{2^{n+1}}.$$

The annuli $\bar{\Delta}_\nu, \bar{\Delta}_{\nu+1}, \ldots$ overlap and have a union equal to

$$E : 0 < |z| \leqslant \frac{R}{2^{\nu+1}}.$$

Therefore $|f(z)| > M$ for all $z \in E$, i.e.,

$$\lim_{z \to 0} f(z) = \infty,$$

which is impossible, since $z = 0$ is an essential singular point of $f(z)$. This contradiction proves the lemma.

We are now in a position to prove the following refinement of Theorem 9.16:

THEOREM 9.17 (*Julia's theorem*). *Let $f(z)$ be analytic on $0 < |z| < R$, with an essential singular point at $z = 0$. Then there is at least one ray L*

emanating from $z = 0$ such that $f(z)$ takes every finite value with one possible exception at every point of an infinite sequence converging to $z = 0$ and lying within any given angle with L as its bisector.

Proof. Introducing the annular domains

$$D_0' : \frac{R}{2^4} < |z| < R, \quad D_1' : \frac{R}{2^5} < |z| < \frac{R}{2}, \ldots, \quad D_n' : \frac{R}{2^{n+4}} < |z| < \frac{R}{2^{n+1}}, \ldots,$$

we again construct the sequence of functions (9.23), where $f_n(z)$ takes the same values in D_0' as $f(z)$ takes in D_n'. According to the lemma, $\{f_n(z)\}$ cannot be normal on D_0'. Therefore D_0' contains at least one point ζ such that $\{f_n(z)\}$ fails to be normal in every neighborhood of ζ, since otherwise, by the argument on p. I.416, $\{f_n(z)\}$ would be normal on D_0'. Let $\mathcal{N}(\zeta; \varepsilon)$ be an ε-neighborhood of ζ contained in D_0'. Since $\{f_n(z)\}$ is not normal on $\mathcal{N}(\zeta; \varepsilon)$, the terms of $\{f_n(z)\}$ take every value in $\mathcal{N}(\zeta; \varepsilon)$ with one possible exception. More exactly, given any A with one possible exception, there exist functions $f_{n_k}(z)$ with arbitrarily large index n_k taking the value A at a point $z_k \in \mathcal{N}(\zeta; \varepsilon)$. Observing that

$$f_{n_k}(z) = f\left(\frac{z}{2^{n_k}}\right)$$

and denoting $z_k/2^{n_k}$ by ζ_k, we have

$$f(\zeta_k) = A.$$

The sequence $\{\zeta_k\}$ obviously converges to the point $z = 0$ while staying inside the angle bounded by the rays drawn from $z = 0$ tangent to the circle $|z - \zeta| = \varepsilon$. The theorem now follows from the fact that ζ is fixed and ε is arbitrarily small.

COROLLARY. *Let $f(z)$ be an entire transcendental function. Then there is at least one nonzero finite point ζ such that the sequence $\{f(2^n z)\}$ fails to be normal in every neighborhood of ζ. Let L be the ray emanating from the origin and passing through ζ. Then $f(z)$ takes every finite value, with one possible exception, at every point of an infinite sequence converging to $z = \infty$ and lying within any given angle with L as its bisector.*

Proof. Use the transformation $Z = 1/z$ carrying the essential singular point $z = \infty$ into the origin.

Remark. Picard's second theorem says nothing about the distribution of A-points of $f(z)$ in a neighborhood of an essential singular point, whereas Julia's theorem asserts that given any finite A, with one possible exception, $f(z)$ has infinitely many A-points near a certain ray, called a *Julia direction.* There may be several or even infinitely many Julia directions (see Problem 9.17).

Example. The function e^z has two Julia directions, namely the positive imaginary axis and the negative imaginary axis. In fact, the roots of the equation $e^z = A$ are of the form

$$z_k = \operatorname{Ln} A = \ln |A| + i \arg A + 2k\pi i.$$

If k is positive, the tangent of the angle between the vector z_k and the positive imaginary axis equals

$$\frac{\ln |A|}{\arg A + 2k\pi}$$

and obviously approaches zero as $k \to \infty$. It follows that starting from some value of k, the points z_k all lie within any fixed angle with the positive imaginary axis as its bisector. A similar result is obtained for $k < 0$ and the negative imaginary axis. It should be noted that the sequence $\{e^{2^n z}\}$ fails to be normal in any neighborhood of every point of the imaginary axis. In fact, $|e^{2^n z}| = 1$ at every such point, while $\{e^{2^n z}\}$ approaches ∞ at every point of the right half-plane and 0 at every point of the left half-plane. Therefore no point of the imaginary axis has a neighborhood in which any subsequence $\{e^{2^{n_k} z}\}$ is uniformly convergent.

PROBLEMS

9.1. Let $\{G_1, f_1\}$ and $\{G_2, f_2\}$ be two elements whose domains are disjoint but share an accessible Jordan boundary arc δ, where δ is open and rectifiable.[23] Suppose $f_1(z)$ is analytic on G_1 and continuous on $G_1 \cup \delta$, while $f_2(z)$ is analytic on G_2 and continuous on $G_2 \cup \delta$. Prove that if $f_1(z)$ and $f_2(z)$ coincide on δ, then there exists a function $\varphi(z)$ analytic on $D = G_1 \cup \delta \cup G_2$ such that

$$\varphi(z) = \begin{cases} f_1(z) & \text{if } z \in G_1, \\ f_1(z) = f_2(z) & \text{if } z \in \delta, \\ f_2(z) & \text{if } z \in G_2. \end{cases}$$

Hint. Use Theorem 3.10.

9.2. Use the symmetry principle to show that every point of the segment BC in Figure 8.4(a), p. 312 is a regular point of the function $f(z)$ constructed in the answer to Problem 8.19.

9.3. Use the symmetry principle to give another proof of Theorem 1.1.

9.4. It will be recalled from Problem 6.5 that $z = sn\, w$ is the inverse of the elliptic integral

$$w = f(z) = \int_0^z \frac{dt}{\sqrt{(1 - t^2)(1 - k^2 t^2)}}.$$

---

[23] By an *open* arc we mean an arc without its end points (rectifiability of an open arc is defined in the obvious way, i.e., by including the end points).

Using the symmetry principle, prove that sn w is meromorphic and doubly periodic with fundamental periods $4K$ and $2iK'$, and then locate the zeros and poles of sn w.

9.5. Use the symmetry principle to map the domain between the branches of the hyperbola

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

onto the upper half-plane.

Hint. See Sec. I.51.

Ans.

$$w = \left(\frac{z + \sqrt{z^2 - c^2}}{ce^{i\theta}}\right)^{\frac{\pi}{\pi - 2\theta}}, \qquad \text{where } c = \sqrt{a^2 + b^2},\ \theta = \arcsin\frac{a}{c}.$$

9.6. Verify Theorems 9.8 and 9.10 by elementary geometry.

9.7. Map the upper half-plane onto the domains shown in Figures 9.4(a)–9.4(d).

9.8. Prove that a necessary condition for a function mapping a bounded polygon onto the upper half-plane to be single-valued is that the polygon have no more than four sides.

9.9. Find the domains in the w-plane onto which the function

$$w = f(z) = \int_0^z \frac{t - \lambda}{\sqrt{t(t-1)}}\,dt$$
$$= \sqrt{z(z-1)} + (1 - 2\lambda)\left[\ln\left(\sqrt{z} + \sqrt{z-1}\right) - \frac{i\pi}{2}\right]$$

maps the upper half-plane $\operatorname{Im} z > 0$ if

a) $\lambda < 0$; b) $0 < \lambda < \frac{1}{2}$; c) $\lambda = \frac{1}{2}$; d) $\frac{1}{2} < \lambda < 1$; e) $\lambda > 1$.

9.10. Find the function $w = f(z)$ mapping the disk $|z| < 1$ onto the exterior of regular n-gon with its center at the origin and one vertex at the point $w = 1$, given that $f(0) = \infty$, $f(x) > 0$ $(0 < x < 1)$.

9.11. The process of continuing $\chi(z)$ into the unit disk described in Sec. 49 can also be carried out by applying the symmetry principle to the polygonal domains into which $\chi(z)$ is continued at each stage (rather than to the smaller triangular domains). Describe this process, starting from the triangle ABC.

Ans. The first step again gives the hexagon $ADBECF$, but subsequent steps give a $6 + 4 \cdot 6 = 30$-gon, a $30 + 28 \cdot 30 = 870$-gon, etc.

9.12. Prove that the unit circle $|z| = 1$ is the natural boundary of the function $\chi(z)$ figuring in Theorem 9.13.

9.13. Use Montel's normality criterion (Theorem 9.15) to deduce Harnack's theorem (Theorem II.5.12).

9.14. Prove that the inverse of an entire function $f(z)$ is itself entire if and only if $f(z) = az + b$.

9.15. Give an example of an entire transcendental function $f(z)$ approaching infinity along every ray emanating from the origin. Reconcile this with the fact that $f(z)$ has an essential singular point at infinity.

Ans. $f(z) = e^z + z$.

9.16. Find the Julia directions of the functions $\tan z$ and $\sin(1/z)$.

9.17. Prove that e^{e^z} has infinitely many Julia directions.

BIBLIOGRAPHY[1]

Ahlfors, L. V. and L. Sario, *Riemann Surfaces*, Princeton University Press, Princeton, N. J. (1960).

Behnke, H. and F. Sommer, *Theorie der Analytischen Funktionen einer Komplexen Veränderlichen*, second edition, Springer-Verlag, Berlin (1962).

Bergmann, S., *The Kernel Function and Conformal Mapping*, American Mathematical Society, New York (1950).

Betz, A., *Konforme Abbildung*, second edition, Springer-Verlag, Berlin (1964).

Bieberbach, L., *Lehrbuch der Funktionentheorie* (in two volumes), Chelsea Publishing Co., New York (1945).

Bieberbach, L., *Conformal Mapping* (translated by F. Steinhardt), Chelsea Publishing Co., New York (1953).

Bieberbach, L., *Analytische Fortsetzung*, Springer-Verlag, Berlin (1955).

Bliss, G. A., *Algebraic Functions*, American Mathematical Society, New York (1933).

Carathéodory, C., *Conformal Representation*, second edition, Cambridge University Press, London (1952).

Chevalley, C., *Introduction to the Theory of Algebraic Functions of One Variable*, American Mathematical Society, New York (1951).

Courant, R., *Dirichlet's Principle, Conformal Mapping, and Minimal Surfaces*, with an appendix by M. Schiffer, Interscience Publishers, Inc., New York (1950).

Davis, P. J., *Interpolation and Approximation*, Blaisdell Publishing Co., New York (1963).

Dinghas, A., *Vorlesungen über Funktionentheorie*, Springer-Verlag, Berlin (1961).

Eichler, M., *Einführung in die Theorie der Algebraischen Zahlen und Funktionen*, Birkhäuser Verlag, Basel (1963).

[1] For a list of general books in English on complex variable theory (many of which are germane to the topics treated here), see the Bibliography to Volume I.

Ford, L. R., *Automorphic Functions*, second edition, Chelsea Publishing Co., New York (1951).

Gattegno, C. and A. Ostrowski, *Représentation Conforme à la Frontière*; *Domaines Généraux*, Mémorial des Sciences Mathématiques, Fascicule 109; *Représentation Conforme à la Frontière*; *Domaines Particuliers*, ibid., Fascicule 110, Gauthier-Villars, Paris (1949).

Golubev, V. V., *Vorlesungen über Differentialgleichungen im Komplexen*, VEB Deutscher Verlag der Wissenschaften, Berlin (1958).

Goluzin, G. M., *Geometrische Funktionentheorie*, VEB Deutscher Verlag der Wissenschaften, Berlin (1957).

Hayman, W. K., *Multivalent Functions*, Cambridge University Press, London (1958).

Heins, M., *Selected Topics in the Classical Theory of Functions of a Complex Variable*, Holt, Rinehart and Winston, New York (1962).

Hornich, H., *Lehrbuch der Funktionentheorie*, Springer-Verlag, Vienna (1950).

Hurwitz, A. and R. Courant, *Funktionentheorie*, fourth edition, with an appendix by H. Röhrl, Springer-Verlag, Berlin (1964).

Jenkins, J. A., *Univalent Functions and Conformal Mapping*, second edition, Springer-Verlag, Berlin (1965).

Julia, G., *Principes Géometriques d'Analyse*, Gauthier-Villars, Paris, *Volume I* (1930), *Volume II* (1932).

Julia, G., *Leçons sur la Réprésentation Conforme des Aires Simplement Connexes*, Gauthier-Villars, Paris (1931).

Julia, G., *Leçons sur la Réprésentation Conforme des Aires Multiplement Connexes*, Gauthier-Villars, Paris (1934).

Kober, H., *Dictionary of Conformal Representations*, second edition, Dover Publications, New York (1957).

Landau, E., *Darstellung und Begründung Einiger Neuerer Ergebnisse der Funktionentheorie*, second edition, Chelsea Publishing Co., New York (1946).

Lavrentev, M. A., *Sur les Fonctions d'une Variable Complexe Représentables par des Séries de Polynomes*, Actualités Scientifiques et Industrielles No. 441, Hermann et Cie., Paris (1936).

Montel, P., *Leçons sur les Séries de Polynomes à une Variable Complexe*, Gauthier-Villars, Paris (1910).

Montel, P., *Leçons sur les Familles Normales de Fonctions Analytiques et Leurs Applications*, Gauthier-Villars, Paris (1927).

Montel, P., *Leçons sur les Fonctions Univalentes ou Multivalentes*, Gauthier-Villars, Paris (1933).

Nehari, Z., *Conformal Mapping*, McGraw-Hill Book Co., New York (1952).

Nevanlinna, R., *Uniformisierung*, Springer-Verlag, Berlin (1953).

Osgood, W. F., *Lehrbuch der Funktionentheorie*, B. G. Teubner, Leipzig, *Volume I* (1928), *Volume II, Part 1* (1929), *Volume II, Part 2* (1932).

Pfluger, A., *Theorie der Riemannschen Flächen*, Springer-Verlag, Berlin (1957).

Privalov, I. I., *Einführung in die Funktionentheorie, Volumes II, III*, B. G. Teubner, Leipzig (1959).

Sario, L. and K. Noshiro, *Value Distribution Theory* (in collaboration with T. Kuroda, K. Matsumoto and M. Nakai), D. Van Nostrand Co., Inc., Princeton, N. J. (1966).

Schaeffer, A. C. and D. C. Spencer, *Coefficient Regions for Schlicht Functions*, with a chapter by A. Grad, American Mathematical Society, New York (1950).

Sewell, W. E., *Degree of Approximation by Polynomials in the Complex Domain*, Princeton University Press, Princeton, N.J. (1942).

Springer, G., *Introduction to Riemann Surfaces*, Addison-Wesley Publishing Co., Inc., Reading, Mass. (1957).

Stoïlow, S., *Leçons sur les Principes Topologiques de la Théorie des Fonctions Analytiques*, second edition, Gauthier-Villars, Paris (1956).

Tricomi, F., *Elliptische Funktionen*, Akademische Verlagsgesellschaft, Geest & Portig K.-G., Leipzig, (1948).

Tsuji, M., *Potential Theory in Modern Function Theory*, Maruzen Co., Ltd., Tokyo (1959).

Von Koppenfels, W. and F. Stallman, *Praxis der Konformen Abbildung*, Springer-Verlag, Berlin (1959).

Walsh, J. L., *Interpolation and Approximation by Rational Functions in the Complex Domain*, third edition, American Mathematical Society, Providence, R. I. (1960).

Weyl, H., *The Concept of a Riemann Surface* (translated by G. R. MacLane), third edition, Addison-Wesley Publishing Co., Inc., Reading, Mass. (1955).

INDEX

A

Abstract Riemann surface(s), 221–227
  analyticity on a domain of, 223
  analyticity at a point of, 222
  angle between two Jordan curves on, 225
  conformal mapping of, 225–227
  equivalent, 223
  nonequivalent, 224
Accessible arc, 71
  domain adjacent to, 71
Accessible boundary points, 52–57
  images of, 54
Adjacent domains, 59, 60, 71
Aleksandrov, P. S., 231
Algebraic addition theorem, 152
Algebraic branch point, 249, 280
Algebraic equation:
  irreducible, 297
  local solutions of, 298
Analytic configuration, 286–297
  elements of, 286
    neighborhoods of, 287
  as a Riemann surface, 294–297
  as a topological surface, 287–294
Analytic continuation, 257 ff.
  direct, 257
  of an element along a curve, 266 ff.
  in a star, 272–275
  "wide-sense," 315
Analytic curve, 241
  regular, 242
Algebraic function(s), 297–308
  Riemann surface of, 308
    topological model of, 308
Approximation in the mean, 116–121
  definition of, 116
Approximation on closed domains, 97–100
Approximation on continua, 100–103
Area theorem, 14
Associated sigma functions, 184, 194–195

B

Barycentric subdivision, 234
Bazilevich, I. E., 26
Bernstein, S. N., 112
Bernstein's lemma, 112
Bernstein's theorem, 114
Bessel's inequality, 123
Bieberbach, L., 26
Bieberbach's conjecture, 26
Birkhoff, G., 170, 183, 227, 258, 305
Boundary correspondence theorem, 70
Boundary element (*see* Prime end)
Branch point(s), 240, 249, 279 ff.
  algebraic, 249, 280
  logarithmic, 281–282
  transcendental, 281–282
Branched element, 285

C

Carathéodory, C., 33, 57
Carathéodory domain, 38
Carathéodory's mapping theorem, 33

Cauchy's integral formula, generalization of, 99
Cauchy's integral theorem, generalization of, 99
Chain of domains, 62
Chain of elements, 258
Chain of triangles, 229
  coherently oriented, 231
Chebyshev polynomials, 106, 111, 112, 131
Circular element, 263
  center of, 264
  disk of, 264
  radius of, 264
  series of, 264
Circular image, 107
Closed circular arcs, 58
  complementary, 58
  domains adjacent to, 59
Coherently oriented chain of triangles, 231
Coherently oriented triangles, 231
Complementary modulus, 204
Complete analytic function, 259
  derivative of, 309
  domain of existence of, 259
  elements of, 259
  single-valued branches of, 259
  values of, 259
Conformal mapping(s), 3–79
  of annular domains, 3–8
  boundary behavior of, 66–77
  of sequences of simply connected domains, 32–41
  of simply connected domains, 8 ff.
  uniqueness theorem for, 13
Conformal radius, 12
Conformal structure, 226
Congruent points, 147
Conical pendulum, 193
Convergence in the mean, 116
Convex curve, 20, 29
Convex domain, 29
Convex univalent function, 29
Coterminal arc, 75
Countable base, 218
Critical element, 285
Curvilinear half-interval(s), 42–52
  identical, 42
  initial point of, 42
  limit set of, 42

D

Dense set, 255
Direct analytic continuation, 257
Discriminant, 170
Distortion theorem, 17
Domain of existence, 259
Domains of univalence, 249
Doubly periodic function, 138
Dzhrbashyan, M. M., 119

E

Element(s), 257
  analytic continuation of, 258
  branched, 285
  chain of, 258
  circular (*see* Circular element)
  continuation of, along a curve, 266
  critical, 285
  direct analytic continuation of, 257
  domain of, 257
  equality of, 257
  generalized (*see* Generalized element)
  regular, 285
  subordinate, 260
  unbranched, 285
Elliptic function(s), 138, 146 ff.
  A-points of, 150
  order of, 148, 150
  poles of, 148
  representation of, 181–183
  residues of, 148
Elliptic integral(s), 171
  of the first kind, 171
  inversion of, 172–178
Equicontinuous sequence of functions, 73
Extension of homeomorphisms, 233

F

Faber polynomials, 104–112
  expansion in series of, 108
  generating function of, 107
  uniqueness property of, 110
Fourier coefficients, 122
Fourier series, 122
Function element (*see* Element)

Function meromorphic on a domain, 87
Function meromorphic on a Riemann surface, 313
Function of bounded characteristic, 130
Function of exponential type, 143
Fundamental period(s), 137, 138
Fundamental period parallelogram, 138
  sharpened definition of, 147

G

Generalized element(s), 285
  analytic continuation of, 286
  canonical form of, 285
  center of, 285
  direct analytic continuation of, 285
  disk of, 285
  radius of, 285
  subordinate, 285
Generalized polygonal curve, 324
Goluzin, G. M., 8, 20, 26, 130

H

Half-open curvilinear interval (*see* Curvilinear half-interval)
Hausdorff space, 217 ff.
  connected subset of, 218
  interior point of a subset of, 218
  limit point of a subset of, 218
  open subset of, 218
Hayman, W. K., 26
H-intervals, 59
  complementary, 59
  domains adjacent to, 60
  end points of, 59
  sequence of, 60
    nested, 60
    shrinking, 60

I

Increasing sequence of open sets approximating an open set, 81
Inner area, 23
Interior mappings, 232–237
Inversion of elliptic integrals, 172–178
Irreducible algebraic equation, 297

J

Jacobi's elliptic functions, 197 ff.
  complementary modulus of, 204
  derivatives of, 198
  fundamental periods of, 197
  modulus of, 198
  poles of, 197
  as ratios of theta functions, 205
  relation to Weierstrass' elliptic function, 198
  zeros of, 197
Jacobi's identity, 206
Jacobi's theta functions, 199 ff.
  behavior under changes of argument, 203
  infinite product expansions of, 208–211
  partial differential equation for, 203
Jenkins, J. A., 26
Jordan arc, 42
  accessible, 71
  coterminal, 75
Jordan domain, 38, 70
Jordan half-interval, 42
Julia directions, 346–347
Julia's theorem, 345

K

Keldysh, M. V., 8, 100, 119, 129
Keldysh's theorem, 100
Kernel (of a sequence of domains), 32
Komm, H., 231
Korovkin, P. P., 129

L

Lambda functions, 195–196
Lattice, 152
  basis of, 152
  reducible, 152
Lavrentev, M. A., 100, 103, 129
Lavrentev's theorem, 103
Lebedev, N. A., 129
Legendre's relations, 179
Limit set, 42
Local uniformizing parameter, 221
Locally analytic functions, 80–84
Locally conformal, 226

Locally Euclidean, 218
Logarithmic branch points, 281–282
Lower domain, 72
Luzin-Privalov boundary uniqueness theorem, 319

M

MacLane, S., 170, 183, 227, 258, 305
Many-sheeted surface, 240
Markushevich, A. I., 3
Menshov, D. E., 3
Menshov's theorem, 3, 225, 227
Mergelyan, S. N., 103
Meromorphic function:
  on a domain, 87
  on a Riemann surface, 313
Metric space, 254
Mittag-Leffler's expansion, 275
Mittag-Leffler star (*see* Star)
Mittag-Leffler's theorem, 84
  generalization of, 84
Möbius strip, 220–221, 231, 234
Modular function:
  construction of, 336–338
  Riemann surface of, 339
Modulus, 198
  complementary, 204
Monodromy theorem, 269
Montel's theorem, 97
Montel's normality criterion, 340

N

Natural boundary, 261
Nehari, Z., 8, 26
Nevanlinna, R., 130
Nonorientable topological surface, 231
Normal family of analytic functions, 340

O

One-sided topological surface, 231
Open arc, 315
Orientable topological surface, 231
Orthogonal polynomials, 121
Orthonormal polynomials, 121

P

Parallelogram lattice, 137
Parametric neighborhood, 221
Parseval's theorem, 123
Pendulum:
  conical, 193
  simple, 187
  spherical, 186–191
Perfect set, 89
Period, 135
  fundamental, 137
  primitive, 137
Period strip, 141
Periodic entire functions, 141–145
  representation of, 141–142
Periodic function, 135
Periodic meromorphic functions, 135–154
  representation of, 145–146
Picard's first (or little) theorem:
  for entire functions, 339
  for meromorphic functions, 339
Picard's second (or great) theorem, 343
  implication for entire functions, 344
  implication for meromorphic functions, 345
Polynomials orthogonal on a domain, 121
Polynomials orthonormal on a curve, 129
Polynomials orthonormal on a domain, 121 ff.
  Fourier coefficients with respect to, 122
  Fourier series with respect to, 122
Prime end(s), 57–66
  accessible point of, 65
  convergence of a chain of domains to, 62
  convergence of a sequence of points to, 66
  definition of, 60
  distinct, 60
  of the first kind, 65
  of the fourth kind, 65
  inaccessible point of, 65
  of the second kind, 65
  of the third kind, 65
Primitive period, 137
Principal star (*see* Star)
Projection, 240
Prüfer manifold, 228

R

Radial boundary values, 94
Rectilinear star (*see* Star)
Regular analytic arcs, 243
  structure of intersection of, 243
Regular analytic curve, 242
  conformal image of, 242
Regular element, 285
Regular points, 276
Relative topology, 254
Riemann, B., 8
Riemann (covering) surface(s), 237–241
  analyticity at a point of, 238
  branch point on, 240
  closed, 241
    topological model of, 241
  equivalent, 238–240
  examples of, 250–254
  of a meromorphic function, 245–254
  nonequivalent, 240
  topological model of, 240
Riemann surfaces, 217–256
  abstract (*see* Abstract Riemann surfaces)
Riemann's mapping theorem, 8
Rotation theorem, 20
Runge's theorem, 88
  stronger version of, 91

S

Schwarz-Christoffel transformation:
  for bounded polygons, 320
  examples of, 332–336
  mapping unit disk onto polygon, 331
  for polygons with ∞ as a boundary point, 325
  for polygons with ∞ as an interior point, 329
Schwarz's reflection principle, 315
Sequence of simply connected domains:
  conformal mapping of, 32–41
  convergence of, 33
  decreasing, 38
  kernel of, 32
Set dense in itself, 80
Shaginyan, A. L., 119
Sharp estimates, 20, 26
Sheets, 240
Silverman, R. A., 3, 101
Simple pendulum, 187
Simple point, 68
Simply periodic function, 136
Single-valued branches, 259
Singular points, 276 ff.
  distinct, 283
  identical, 283
  isolated, of single-valued character, 278
Smirnov, V. I., 129
Spherical pendulum, 186–191
Square-integrable function, 116
Star, 273
  analytical continuation in, 272–275
Star of triangles, 228
Starlike curve, 29
Starlike domain, 29
Starlike univalent function, 29
Stoïlow, S., 228, 232
Subordinate element, 260, 285
Symmetry principle, 315 ff.

T

Tolstov, G. P., 101
Topological space, 217
  points of, 217
    neighborhoods of, 217
Topological surface(s), 218 ff.
  closed, 218
  compact, 218
  examples of, 218–221
  interior mapping of, 232
  nonorientable, 231
  one-sided, 231
  open, 218
  orientable, 231
  triangulation of, 228
  two-sided, 231
Topological triangle(s), 227
  chain of, 229
    coherently oriented, 231
  chained, 229
  coherently oriented, 231
  oriented, 227
  sides of, 227
  star of, 228
  vertices of, 227

Torus, 219
Transcendental branch point, 280
Triangulable manifold, 228
Triangulation, 228
Tricomi, F., 169
Trigonometric polynomials, 142–145
Two-dimensional manifold, 218
  triangulable, 228
    triangulation of, 228
Two-sided topological surface, 231
Typically real function, 31

U

Univalent function(s), 14–26
  convex, 29
  starlike, 29
Universal function, 92
Upper domain, 72

W

Walsh's theorem, 98
Weierstrass, K., 159, 312
Weierstrass' elliptic function(s), 155 ff.
  addition theorem for, 185
  definition of, 159
  differential equation for, 169
  duplication formula for, 193
  invariants of, 169
  partial fraction expansion of, 211
  properties of, 160–168
  Riemann surface of, 253
Weierstrass' sigma function, 180
Weierstrass' theorem, 86
  generalization of, 86
Weierstrass' zeta function, 178
  partial fraction expansion of, 211
  quasi-periodicity of, 179
Whittaker, E. T., 187
Widder, D. V., 5

CONTENTS OF VOLUME I

PART 1. BASIC CONCEPTS

Chap. 1. Introduction
Chap. 2. Complex Numbers
Chap. 3. Sets and Functions. Limits and Continuity
Chap. 4. Connectedness, Curves and Domains
Chap. 5. Infinity and Sterographic Projection
Chap. 6. Homeomorphisms

PART 2. DIFFERENTIATION. ELEMENTARY FUNCTIONS

Chap. 7. Differentiation and the Cauchy-Riemann Equations
Chap. 8. Geometric Interpretation of the Derivative. Conformal Mapping
Chap. 9. Elementary Entire Functions
Chap. 10. Elementary Meromorphic Functions
Chap. 11. Elementary Multiple-Valued Functions

PART 3. INTEGRATION. POWER SERIES

Chap. 12. Rectifiable Curves. Complex Integrals
Chap. 13. Cauchy's Integral Theorem
Chap. 14. Cauchy's Integral and Related Topics
Chap. 15. Uniform Convergence. Infinite Products
Chap. 16. Power Series: Rudiments
Chap. 17. Power Series: Ramifications
Chap. 18. Methods for Expanding Functions in Taylor Series

CONTENTS OF VOLUME II

PART 1. LAURENT SERIES. CALCULUS OF RESIDUES

Chap. 1. Laurent Series. Isolated Singular Points
Chap. 2. The Calculus of Residues and Its Applications
Chap. 3. Inverse and Implicit Functions
Chap. 4. Univalent Functions

PART 2. HARMONIC AND SUB-HARMONIC FUNCTIONS

Chap. 5. Basic Properties of Harmonic Functions
Chap. 6. Applications to Fluid Dynamics
Chap. 7. Subharmonic Functions
Chap. 8. The Poisson-Jensen Formula and Related Topics

PART 3. ENTIRE AND MEROMORPHIC FUNCTIONS

Chap. 9. Basic Properties of Entire Functions
Chap. 10. Infinite Product and Partial Fraction Expansions

91382

SELECTED RUSSIAN PUBLICATIONS IN THE MATHEMATICAL SCIENCES

RICHARD A. SILVERMAN, *Editor*

Formerly of the *Courant Institute, New York University*

AN INTRODUCTION TO THE THEORY OF LINEAR SPACES
by G. E. SHILOV
Published 1961 — *310 pages*

FOURIER SERIES
by G. P. TOLSTOV
Published 1962 — *336 pages*

EXTRACTION OF SIGNALS FROM NOISE
by L. A. WAINSTEIN and V. D. ZUBAKOV
Published 1962 — *382 pages*

AN INTRODUCTION TO THE THEORY OF STATIONARY RANDOM FUNCTIONS
by A. M. YAGLOM
Published 1962 — *235 pages*

CALCULUS OF VARIATIONS
by I. M. GELFAND and S. V. FOMIN
Published 1963 — *232 pages*

METHODS OF QUANTUM FIELD THEORY IN STATISTICAL PHYSICS
by A. A. ABRIKOSOV, L. P. GORKOV and I. E. DZYALOSHINSKI
Published 1963 — *352 pages*

SPECIAL FUNCTIONS AND THEIR APPLICATIONS
by N. N. LEBEDEV
Published 1965 — *308 pages*

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE, Volume I
by A. I. MARKUSHEVICH
Published 1965 — *459 pages*

THEORY OF FUNCTIONS OF A COMPLEX VARIABLE, Volume II
by A. I. MARKUSHEVICH
Published 1965 — *333 pages*

PROBLEMS OF MATHEMATICAL PHYSICS
by N. N. LEBEDEV, I. P. SKALSKAYA and Y. S. UFLYAND
Published 1965 — *429 pages*

INTEGRAL, MEASURE AND DERIVATIVE: A UNIFIED APPROACH
by G. E. SHILOV and B. L. GUREVICH
Published 1966 — *233 pages*

ORDINARY DIFFERENTIAL EQUATIONS
by I. G. PETROVSKI
Published 1966 — *232 pages*